الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة والنظرية الكلاسيكية. " دراسة سيكومترية "

الدكتورة / غادة خالد عيد
قسم علم النفس التربوي ـ كلية التربية ـ جامعة الكويت

صفحة بيضاء

الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة والنظرية الكلاسيكية. " دراسة سيكومترية "

الدكتورة / غادة خالد عيد

ملخص الدراسة

هدفت هذه الدراسة إلى: ١- الكشف عن الفروق في الدرجات الخام لليقظة العقلية بين الذكور والإناث وبين التخصصات العلمية و الأدبية.٢- الكشف عن الفروق في الدرجات الحقيقية لليقظة العقلية والمقدرة بنموذج راش بين الذكور والإناث، وبين التخصصات الأدبية والعلمية. ٣- الكشف عن الفروق في الدرجات الحقيقية لليقظة العقلية المقدرة بالنماذج الكلاسيكية بين الذكور والإناث، وبين التخصصات العلمية والأدبية. ٤- الكشف عن الفروق بين الدرجة الحقيقية المقدرة بنموذج راش وتلك المقدرة بالنظرية الكلاسيكية. وتكونت عينة الدراسة من ٢٥٠ طالب وطالبة من طلاب وطالبات كلية التربية ، بقسميها العلمي (رياضيات، كيمياء، فيزياء) (١٥٠طالب وطالبة) والأدبي (اللغة العربية، اللغة الإنجليزية) (١٠٠طالب وطالبة) وطبق عليهم اختبار اليقظة العقلية كمقياس للذكاء العام. وتم التعويض في المعادلة ت = م + رأ (س – م). حيث "ت" تعني الدرجة الحقيقية، رأ معامل الثبات، م متوسط الدرجات، س الدرجة الخام ، للحصول على الدرجات الحقيقية المناظرة للدرجات الخام باستخدام النظرية الكلاسيكية. واستخدمت طريقة PROX للحصول على القدرات الحقيقية للطلاب مقدرة باللوجيت من نموذج راش كأحد نماذج السمات الكامنة. وباستخدام اختبار "ت" T-Test بينت النتائج: وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للطلاب والطالبات في مقياس اليقظة العقلية، وكذلك وجود فروق ذات دلالة إحصائية بين متوسطي التخصصات الأدبية والتخصصات العلمية في مقياس اليقظة العقلية، وعدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (نموذج راش) للطلاب والطالبات على مقياس اليقظة العقلية. وعدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية للتخصصات العلمية والأدبية على مقياس اليقظة العقلية، عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية (التقليدية)

~~للطلاب والطالبات على مقياس اليقظة العقلية.عدم وجود فروق ذات دلالة إحصائية بين متوسطي~~
الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية للتخصصات العلمية والتخصصات الأدبية في مقياس اليقظة العقلية، وعدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بنموذج راش وتلك المقدرة بالمدخل الكلاسيكي لمقياس اليقظة العقلية.

True Score Estimated by Latent Trait Theory and Classical Test Theory "A Psychometric Study"

Dr. Ghada Khalid Eid

Abstract

The purpose of this study is to investigate: 1- The difference between male and female in intelligence obtained score and the difference between scientific specializations and literacy specializations in ontelligence obtained score. 2- The difference between male and female in intelligence true score estimated by rasch model and the difference between scientific specializations and literacy specializations in intelligence true score estimated by rasch model.3- The difference between male and female in intelligence true score estimated by classical test theory and the difference between scientific specializations and literacy specializations in intelligence true score estimated by classical test theory. 4- The difference between true score estimated by latent trait theory (rasch model) and classical test theory. General intelligence test was applied on (250) subjects divided into (136) male and (114) female according to sex and (150) scientific specializations, (100) literacy specializations according to specialization. The equation t = m + r(s-m) was used to estimate true score from classical test theory , and prox method was used to estimate latent trait to rasch model . Independent sample " t " test was used to statistical treatment. Results indicated that: 1- There was a significant difference between male and female in intelligence obtained score and the same to difference between scientific specializations and literacy specializations in intelligence obtained score. 2-there was no significant difference between male and female in intelligence true score estimated by rasch model and the same to difference between scientific specializations and literacy specializations in intelligence true score estimated by rasch model.3- there was no significant difference between male and female in intelligence true score estimated by classical test theory and the same to difference between scientific specializations and literacy specializations in intelligence true score estimated by Classical Test

theory. 4-There was no significant differences between true scores estimated by classical test theory and those estimated by rasch as a model of latent trait theory.

مقدمــــــــــة:

يعد القياس والتقويم التربوي أحد المكونات الرئيسة للمنظومة التربوية، وأكثرها تأثيراً في تقدمها وازدهارها، ويسهم تطويرها إسهاماً إيجابياً في إصلاح وتطوير بقية تلك المكونات. لذلك فقد حظى مجال تطوير نظم وأساليب القياس والتقويم باهتمام كبير من قبل كل المهتمين بتطوير النظم التعليمية استناداً إلى ما أكدته البحوث والدراسات التربوية المعاصرة من الأهمية البالغة لعمليات وأساليب القياس المتطورة في توجيه مسار العمل التربوي والنهوض به، وتحديد مدى تحقيق النظام التعليمي لأهدافه المرجوة. وتعد الاختبارات والمقاييس النفسية أحد الوسائل الهامة والضرورية لعمليات التقويم التي تعتمد عليها المؤسسات التعليمية والتربوية للتعرف على مقدار التقدم في تحصيل الطلاب ومدى تحقيقهم للأهداف التعليمية، لذلك يمكن اعتبارها أساسا هاما في صنع بعض القرارات المصيرية لكل من تطبق عليه هذه الاختبارات. وبالتالي فإن دقة وسلامة تلك القرارات تتوقف على مدى جودة وكفاءة الاختبارات المستخدمة في عملية التقويم والأساليب المستخدمة في تحليل وتفسير نتائجها. (علام، ١٩٨٧: ٢١).

حيث يعتمد تقويم الطلاب على تحليل البيانات التي يتم الحصول عليها عن طريق الاختبارات وأدوات القياس المختلفة، كما ويعتمد على كافة البيانات الأخرى التي تساعد على معرفة التغيرات التي تطرأ على نمو الطالب في جميع جوانب شخصيته. ومن هنا تظهر أهمية سلامة البيانات التي تم التقويم في ضوئها، وبالتالي التأكد من كفاءة الأدوات التي تستخدم في جمع هذه البيانات. (القرشي، ١٩٨٦: ٤٠٣) وبالرغم من الإيمان المسبق بوجود خطأ في تقدير الدرجة التي يحصل عليها الطالب في اختبار ما بالزيادة أو النقصان، فإنه يجب ألا يصل الخطأ إلى درجة تصبح عندها هذه الدرجة مضللة للطالب والمدرس وكل من يهمه أمرها. فقد يشعر طالب بأن الدرجة التي حصل عليها تبخس تحصيله الحقيقي في حين يشعر آخر بأنها

تبالغ في تحصيله الحقيقي، إلا أنه بحكم رغبته في درجة عالية قد يبدي اتجاهها إيجابياً نحو المقرر ومدرسه، ولكن سرعان ما يكتشف أن درجته مجرد رقم مضلل عندما يخضع لمحك معين، وعندها قد يبدأ اتجاهه الإيجابي بالتحول . (عودة، وحوامدة، ١٩٩٦: ٤٣٣)

فالدرجة التي يحصل عليها الطالب في أي اختبار تحصيلي أو أي قياس عقلي تتكون من جزئين هما: جزء حقيقي "ت" وجزء يرجع إلى خطأ القياس "خ" وبفرض أن هذه الدرجة التي يحصل عليها الطالب في اختبار ما هي "س" فإن: س = ت + خ

ومن هذه العلاقة تنضح أهمية معرفتنا إلى أي مدى تكون الدرجة الخام أو الخام (س) خالية من أخطاء وشوائب القياس ليكون اتخاذ القرار بشأن هذا الطالب وتوجيهه توجيهاً سليماً ودقيقا. (سعيد، ٢٠٠٠ : ٥٣)

أي أن الدرجة التي نحصل عليها من تطبيق الاختبار يطلق عليها الدرجة الخام Obtained Score، وهذه الدرجة خليط من الدرجة الحقيقية True Score وخطأ القياس Measurement Error وبطرح درجة خطأ القياس من الدرجة الخام نحصل على الدرجة الحقيقية، أي أن الدرجة الحقيقية = الدرجة الخام ـ خطأ القياس. وقد يكون خطأ القياس موجباً أو سالباً أو صفراً، وبالتالي فإن الدرجة الخام التي نحصل عليها من أي اختبار أو أي قياس عقلي قد تكون أكبر من أو أقل من الدرجة الحقيقية أو مساوية لها. (Anthony, 1983: 38)

ولذلك فقد يحصل بعض الأفراد على درجات مرتفعة في حين أن مستواهم الحقيقي أقل من ذلك، وقد يحدث العكس حيث يحصل بعض الأفراد على درجات منخفضة في حين أن مستواهم الحقيقي أو قدراتهم الحقيقية أعلى من أو تفوق الدرجة الخام التي حصلوا عليها، أما الحالة المثالية لدقة الاختبار وكفاءته فإن الدرجة الخام

يجب أن تكون مساوية للدرجة الحقيقية ومن هنا تأتي أهمية معرفة الدرجات الحقيقية المناظرة للدرجات الخام لأي اختبار أو قياس عقلي.

وتميز أدبيات القياس التربوية والنفسي بين مدخلين رئيسيين في تقييم وبناء الاختبارات والمقاييس وتحليل البيانات المستمدة منها، وهذين المدخلين هما المدخل الكلاسيكي أو التقليدي والمدخل المعاصر. والمدخل الأول يعتمد على النظرية الكلاسيكيـــة للاختبـــارات Classical Test Theory وما تنطـــوي عليه من مفاهيم ومبادئ بعضها يتعلق بخصائص مفردات الاختبار (الصعوبة ، التمييز) والبعض الآخر يتعلق بخصائص الاختبار ككل (الصدق، الثبات، والمعايير). أما المدخل الثاني فيعتمد على النظرية المعاصرة للاختبارات Modern Test Theory والتي تعرف باسم نظرية السمات الكامنة (Latent Trait Theory (LTT، والتي يفضل علماء القياس من أمثال لورد Lord تسميتها نظرية الاستجابة على المفردة الاختبارية(Item Response Theory (IRT لاهتمامها بالربط بين استجابة الفرد على مفردة اختبار ما وبين خصائص هذه المفردة. (علام، ١٩٩٠: ١٢٩)

ولكل من المدخلين طرقه الخاصة في تقدير قدرة الفرد الحقيقية من الدرجة الخام obtained score، وهذه الطرق تختلف في المدخلين الكلاسيكي أو التقليدي، والمعاصر أو نظرية السمات الكامنة، ومن هنا نبعت مشكلة الدراسة الحالية في محاولة التعرف على مدى تشابه أو اختلاف الدرجة الحقيقية المقدرة باستخدام المدخل الكلاسيكي وتلك المقدرة باستخدام مدخل السمات الكامنة.

مشكلة الدراسة:

تعتبر نظرية السمات الكامنة من الاتجاهات الحديثة في مجال القياس والتقويم النفسي والتربوي، والتي ظهرت لتعالج الكثير من السلبيات التي وجهت للنظرية الكلاسيكية في القياس. ولقد تعددت وتنوعت الدراسات التي تناولت هذا المدخل ـ السمات الكامنة ـ والملاحظ على هذه الدراسات تركيزها على تحليل بيانات الاختبارات العقلية والمعرفية كما في دراسة علام (١٩٨٥)، ودراسة يوسف

(١٩٩١)، أو استخدام هذه النظرية في بناء الاختبارات التحصيلية كما في دراسة علام (١٩٩٠)، وغير هذه الدراسات الكثير، وعلى الرغم من كثرة هذه الدراسات التي تناولت السمات الكامنة واختلاف أهدافها إلا أنه لا توجد دراسة تناولت المقارنة بين الدرجة الحقيقية True Score المقدرة باستخدام النظرية الحديثة في القياس أو السمات الكامنة، وتلك الدرجة الحقيقة المقدرة من النماذج التقليدية الكلاسيكية، كمحاولة لتقييم مدى كفاءة ودقة هذه النظرية الحديثة في قياس السمات الكامنة أو القدرات الحقيقية، حيث إنه من المفترض ومن خلال الإطار النظري الذي سنعرض له، أن النماذج الحديثة أفضل من المدخل الكلاسيكي في تقدير السمات الكامنة أو الدرجات الحقيقية، وهذه الدراسة محاولة للتحقق من صحة هذا الفرض من خلال الإجابة عن التساؤلات الآتية:

السؤال الأول:

هل تختلف الدرجات الخام Obtained scores للطلاب على مقياس اليقظة العقلية* باختلاف:

- الجنس (طلاب - طالبات).
- التخصص (علمي، أدبي).

السؤال الثاني:

هل تختلف الدرجات الحقيقية True scores للطلاب على مقياس اليقظة العقلية والمقدرة باستخدام نظرية السمات الكامنة (نموذج راش) باختلاف:

* مقياس اليقظة العقلية: هو المقياس الأول من بطارية الاستعداد العقلي للمرحلة الثانوية والجامعية ومن إعداد الغريب (١٩٦٣)، واختارته الباحثة لمناسبته للتحليلات الخاصة بنموذج "راش" كما يعتبر من أكثر مقاييس الذكاء استخداما في البحوث العربية.

ـ الجنس. ـ التخصص.

السؤال الثالث:

هل تختلف الدرجات الحقيقية للطلاب على مقياس اليقظة العقلية والمقدرة باستخدام النظرية الكلاسيكية باختلاف:

ـ الجنس. ـ التخصص.

السؤال الرابع:

هل تختلف الدرجات الحقيقية المقدرة باستخدام نموذج راش عن الدرجات الحقيقية المقدرة بإستخدام النماذج الكلاسيكية؟

أهداف الدراسة:

تهدف الدراسة الحالية إلى:

١ـ الكشف عن الفروق في الدرجات الخام (الدرجات الخام) بين الذكور والإناث وبين التخصصات العلمية والأدبية في مقياس اليقظة العقلية.

٢ـ الكشف عن الفروق في الدرجات الحقيقية المقدرة بنموذج راش بين الذكور والإناث، وبين التخصصات المختلفة في مقياس اليقظة العقلية.

٣ـ الكشف عن الفروق في الدرجات الحقيقية المقدرة بالنماذج الكلاسيكية بين الذكور والإناث، وبين التخصصات العلمية والأدبية في مقياس اليقظة العقلية.

٤ـ الكشف عن الفروق بين الدرجة الحقيقية المقدرة بنموذج راش وتلك المقدرة بالنظرية الكلاسيكية.

أهمية الدراسة:

ترجع أهمية هذه الدراسة إلى الموضوع الذي تتناوله وهو المقارنة بين الدرجات الحقيقية المقدرة باستخدام نموذج راش وتلك المقدرة باستخدام المدخل الكلاسيكي، وهو من الموضوعات الحديثة في مجال القياس والتقويم التربوي والنفسي، وذلك كمحاولة للتأكد من مدى دقة وكفاءة نماذج السمات الكامنة (نموذج راش) فيما وضعت من أجله، كما أن ندرة الدراسات العربية في هذا المجال تزيد

من أهمية هذه الدراسة، وقد تفتح المجال أمام عدد من الباحثين لإثراء مجال القياس والتقويم بالدراسات التجريبية والمرتبطة بهذا الموضوع.

الإطار النظري للدراسة:

نظرا لأن هذه الدراسة تهتم بالدرجة الحقيقية المقدرة من نماذج السمات الكامنة وتلك المقدرة من النظرية الكلاسيكية، فسوف نعرض فيما يلي لمفهوم الدرجة الحقيقية، والطرق المستخدمة في حسابها من خلال عرض نظرية السمات الكامنة ونموذج راش كأحد هذه النماذج والتي سوف تستخدم في معالجة بيانات الدراسة، بعد عرض مفهوم الثبات وكيفية الحصول على الدرجة الحقيقية من خلال قيم معاملات الثبات.

أولاً: مفهوم الدرجة الحقيقية:

كما سبق وأن بينا فإن الدرجة التي نحصل عليها من تطبيق الاختبار تسمى بالدرجة الخام Obtained Score ، وهذه الدرجة خليط من الدرجة الحقيقية True Score وخطأ القياس Measurement Error. وبطرح درجة خطأ القياس من الدرجة الخام نحصل على الدرجة الحقيقية، أي أن الدرجة الحقيقية = الدرجة الخام - درجة خطأ القياس.

وتعبر الدرجة الحقيقية كذلك عن متوسط درجات الفرد الواحد إذا طبق عليه الاختبار عدداً لانهائيا من المرات أو بتطبيق عدد كبير جداً من الاختبارات التي تقيس وظيفة معينة على نفس الفرد، فالدرجة الحقيقية قد نصل إليها عن طريق تطبيق مقياس كامل الدقة. (Paul , 1991: 630)

فحينئذ تعبر الدرجة الحقيقية عن قدرة المفحوص بصدق، أي أنها تقدير حقيقي لما لدى الفرد من قدرة معينة أو صفة محددة. (غنيم ، ١٩٩١ : ٢٨) ولذلك فإن القياس في هذه الحالة يكون بدون أخطاء وتعبر الدرجة على الاختبار عن الدرجة الحقيقية أو القدرة الحقيقية للفرد. (عبده، وعثمان ، ١٩٩٥ : ٢٢٣)

ومن ثم فإن الدرجة الحقيقية (ت) هي جزء من الدرجة التي يحصل عليها الفرد في أي اختبار (س)، وبالتالي فقيمة هذه الدرجة تعتبر مؤشرا لمدى دقة الاختبار في تقدير الخاصية أو الصفة أو السمة التي يقيسها الاختبار، فكلما اقتربت

الدرجة الحقيقية لقدرة الفرد من القيمة المقدرة بالاختبار زادت ثقتنا في هذا الاختبار ونتائجه، وتصل هذه الثقة والدقة ذروتها عندما تعبر الدرجة الخام للفرد عن قدرته الحقيقية، أي عندما س = ت وتصبح درجة خطأ القياس (خ) في هذه الحالة مساوية للصفر (سعيد، ٢٠٠٠: ٥٤). ولذلك ظهرت نماذج السمات الكامنة ومن بين أهدافها الوصول إلى القدرة الحقيقية للفرد من خلال طرقا أكثر دقة وموضوعية من النظرية الكلاسيكية. والتي تعتمد على الثبات في تقدير الدرجة الحقيقية للأفراد المناظرة لدرجاتهم الخام في الاختبارات المختلفة. كما يمكن أيضا تقدير الدرجات الحقيقية المناظرة للدرجات الخام من خلال المدخل الكلاسيكي أو التقليدي في القياس والذي يعتمد بصورة كلية على الثبات في تقدير تلك الدرجات الحقيقية، وذلك على النحو التالي:

ثانياً : العلاقة بين الثبات والدرجة الحقيقية:

يستخدم الثبات في تقييم الاختبارات عن طريق تقدير الدرجات الحقيقية True Sore المناظرة للدرجات الخام وتقدير خطأ القياس بالإضافة إلى تحديد فترات الثقة.(George , 1986: 91)

فالاختبار الذي ليس على درجة مقبولة من الثبات، أو كان خطأ التجريب في القياس كبيراً فإن نتائج هذا الاختبار لا تعبر عن المستويات الحقيقية لقدرة الأفراد، وبالتالي فالقرارات التربوية التي تتخذ بشأن هؤلاء الأفراد سواء بنقلهم إلى مستويات أعلى أم تصنيفهم في مجموعات، تكون مضللة وغير سليمة ويشوبها كثير من الأخطاء. (سعيد، ٢٠٠٠: ٤٢)

فللثبات أثر على جودة القرارات التربوية، فلو افترضنا أن مدرساً ما طبق اختباراً على مجموعة من الطلاب وحدد المستوى ٨٠% لتكون محكاً للنجاح، فإن ثبات الاختبار يجعل من المحتمل أن التلاميذ الذين حصلوا بالفعل على ٨٠% أو أكثر قد يصنفوا على أنهم راسبون، أي يكون مستواهم في هذا الاختبار أقل من ٨٠%، والعكس قد يصنف بعض الطلاب على أنهم ناجحون في حين أن مستوياتهم الحقيقية أقل من محك النجاح المحدد للاختبار. (Anthony , 1983: 389)

وقد توصل " الصياد وستوكر " (El-sayyad and Stoker, 1980) إلى أنه توجد علاقة بين شكل المنحنى المرسوم لمعاملات الثبات المختلفة للاختبار آخذين في الاعتبار درجة القطع، مع شكل التوزيع التكراري لدرجات الطلاب وأشارا إلى أن هذا المنحنى يكون بمثابة صورة معكوسة Mirror Image للتوزيع التكراري. (p. 85)

كما قد يرجع انخفاض قيمة معامل الثبات إلى تباين خطأ القياس، والانخفاض في تباين الدرجات الحقيقية. (Mathalon and others, 1993:179)

فدرجة أي فرد في أي اختبار تنقسم إلى جزئين: جزء جوهري ثابت لا يتأثر بالعوامل الخارجية، ويسمى بالدرجة الحقيقة، وجزء يتأثر بالعوامل الخارجية ويختلف تبعاً لاختلاف هذه العوامل، ويسمى بالدرجة الخاطئة، ومعامل ارتباط تلك الأجزاء الخاطئة يساوي صفراً، ومن ثم فعندما نحسب معامل ارتباط الاختبار بنفسه، فإننا نحصل على قيمة عددية تدل على الجزء الثابت من هذا الاختبار، أي أن الثبات يقيس الجزء الحقيقي من التباين العام للاختبار، وهذا الجزء الحقيقي هو الذي يعطينا القيمة العددية لارتباط الاختبار بنفسه. (السيد، ١٩٧٨: ٣٨٠)

فعندما يطبق الاختبار على مجموعة من الأفراد ثم نعيد تطبيق نفس الاختبار (أو صورة مكافئة له) بعد فترة مناسبة من الزمن على نفس المجموعة فإن معامل الارتباط بين التطبيقين هو ما يعرف بمعامل الثبات، وهذه طريقة من طرق تقدير الثبات تُسمى طريقة الإعادة Test Retest Method.

فإذا افترضنا أن "س" تمثل الاستجابة المقاسة للمفردة، وأن "ت" تمثل الاستجابة الحقيقية المناظرة، فإلى أي حد تتطابق الاستجابة " س" مع نظيرتها "ت" هو ما يعرف بالثبات. وفي هذا التعريف تم وضع لفظة " الاستجابة " بدلاً من " الدرجة " ولفظة " تتطابق " بدلاً من " ترتبط " وذلك لأن الاستجابة للمفردة يمكن أن تكون غير كمية وتتبع تدرجاً معيناً (اسمية مثلاً) كما أن لفظة ترتبط في العادة مقرونة بمعامل ارتباط بيرسون الذي يتطلب أن تكون الاستجابة كمية تنبع من مستوى الفترة، بالإضافة إلى شروط إضافية أخرى، أو ترتيبية حيث يحل معامل ارتباط سبيرمان للرتب محل معامل ارتباط بيرسون بالإضافة إلى أن الارتباط في حده الأعلى تطابق. (الصياد، ١٩٨٨ : ٣٠٩)

ويشير "جونز" (Jones, 1991) إلى أنه إذا كان صدق الاختبار يعبر عن مدى قدرة درجة الفرد في اختبار ما على التبنؤ بدرجته في اختبار آخر، فإن الثبات يشير إلى أي مدى تستطيع درجة هذا الفرد أن تتنبأ بنفسها. (Paul , 1991: 631) وقد يقصد بقدرة الدرجة على التنبؤ بنفسها هو حصول الفرد على نفس الدرجة في قياسات متكررة لنفس الاختبار، ويرتبط هذا التعريف بمفهوم الدرجة الحقيقية كما جاء في تعريف السيد للثبات (السيد ، ١٩٧٨).

كذلك يشير ثبات الاختبار إلى مدى اتساقه مع نفسه في قياس أي جانب يقيسه، أي مدى استقرار الدرجات التي يحصل عليها نفس الأفراد في مرات التطبيق، سواء أعيد التطبيق بنفس الصورة أو بصورة مكافئة لنفس الاختبار. (عبد السلام ،١٩٨٧ : ١٦٣) ويعني هذا الاتساق أو الاستقرار أمرين:

الأول: أن وضع الفرد أو ترتيبه بالنسبة لمجموعته لا يتغير جوهرياً من تطبيق لآخر.

الثاني: أنه لو تكررت عمليات القياس للفرد الواحد لظهرت صفة الاستقرار في درجته في المرات المختلفة. وبذلك يمكن القول بأن ثبات الاختبار يعني ثبات درجة الفرد، ثم ثبات ترتيبه إذا تكرر تطبيق الاختبار (محمد، ١٩٩٦ : ٧٨).

كما يشير ثبات الاختبار إلى عدم وجود خطأ في القياس.(الشرقاوي وآخرون، ١٩٩٦ :٧٧) وإلى دقة الاختبار في القياس أو الملاحظة وعدم تناقضه مع نفسه واتساقه واطراده فيما يزودنا به من معلومات عن سلوك المفحوصين، ويستخدم الثبات بمعناه الواسع ليدل على مدى اعتماد الفروق الفردية في درجات الاختبار على أخطاء الصدفة المتضمنة في القياس. (أبو حطب وآخرون، ١٩٩٧ : ١٠٢)

ويشير معامل الثبات إحصائياً إلى نسبة تباين الدرجة الحقيقية إلى تباين الدرجات الخام، أو أنه مربع معامل الارتباط (معامل التحديد) بين الدرجات الحقيقية والدرجات الخام. (William and Stephen, 1985 : 65)

ومن خلال العرض السابق لمفهوم الثبات فإن الاختبار لا يكون ثابتاً إلا إذا تحقق ما يلي:

١ــ الحصول على نفس النتائج تقريباً إذا أعيد تطبيق الاختبار أو صورة

مكافئة له على نفس المجموعة من الأفراد، بما يدل على أن الاختبار يقيس الأداء الحقيقي للأفراد.

٢ــ ارتفاع نسبة المكون الحقيقي في التباين لدرجات الاختبار أي أن ع٢ح/ع٢ك تكون أعلى ما يمكن، بينما ع٢خ/ع٢ك تكون أقل ما يمكن . وذلك لأن:

التباين الكلي (ع٢ك) = التباين الحقيقي (ع٢ح) + تباين الخطأ (ع٢خ)، ومن ثم فإن:

$$\text{معامل الثبات} = \frac{\text{التباين الحقيقي}}{\text{التباين العام}} = \frac{ع^2_ح}{ع^2_ك}$$

(عبد الرحمن ، ١٩٨٨ : ١٦٣ ـ ١٦٥)

يتضح مما سبق ومن خلال عرض مفهوم الثبات العلاقة الوطيدة و القوية بين مفهوم الثبات للاختبار، والقدرات الحقيقية أو الدرجات الحقيقية للأفراد على هذا الاختبار، حيث إنه كلما ارتفعت قيمة معامل ثبات أي اختبار فإن درجات هذا الاختبار تكاد تكون خالية من أخطاء القياس وتقترب بدرجة كبيرة من القيم الحقيقية لها.وقد وجد أن الاختلاف بين الدرجات الخام (س) والدرجات الحقيقية (ت) دالة لكل من ثبات الدرجات الخام للاختبار (س أ أ) ومتوسط الدرجات الخام (م)، ويمكن الحصول على الدرجات الحقيقية المناظرة للدرجات الخام في أي اختبار من المعادلة التالية: ت = م + س أ أ (س ـ م). (El – Sayyad and stoker , 1980 : 91)

ثالثاً: نظرية السمات الكامنة

مع النمو المتزايد في الوقت الحاضر لتفريد التعليم وقياس نتائج البرامج التعليمية بصورة صادقة وموضوعية ودقيقة، وتمكين المعلمين والمربين من تقدير المستوى الفعلي لأداء الطلاب والاستفادة منه في عملية التوجيه والإرشاد النفسي والتربوي والتعليمي والمهني، فقد تركز الاهتمام وتزايد من جانب علماء القياس وأخصائي بناء الاختبارات والمقاييس النفسية والتربوية في العقدين الأخيرين حول البحث عن وسائل أكثر ملاءمة لتيسير عملية القياس ومقابلة المطالب التعليمية المتطورة، ومواجهة الكثير من المشكلات السيكومترية ومعالجة أوجه القصور

المختلفة التي عجزت عن حلها النظرية السيكومترية التقليدية أو النظرية الكلاسيكية والتي سادت وسيطرت على الفكر التربوي السيكولوجي فترة طويلة من الزمن تزيد عن نصف قرن، ومن أبرز هذه العيوب والمشكلات ما يلي:

١ـ حذف عناصر ووحدات الاختبار غير المميزة بين أفراد العينة، مهما كانت تلك الوحدات أو المفردات ذات قيمة أو ممثلة للأهداف المراد قياسها.

٢ـ الاقتصار على اتخاذ الجماعة كمعيار أوحد لتفسير درجة الاختبار مما يفقدنا البصيرة بموقع الفرد وجماعته التي ينتمي إليها على السواء بالنسبة لأهداف التعلم وأسباب ذلك، وطرق مواجهته.

٣ـ سطوة المنحني الاعتدالي المعياري، فالابتعاد عنه يعتبر انحرافاً سواء في مدى صعوبة أو سهولة المفردات، إذ يؤدي ذلك إلى التواء توزيع اعتدالها. (يوسف، ١٩٩١ : ٤٤٣)

ويعتبر المنحنى الاعتدالي أكثر التوزيعات ملاءمة لفرصة الاختيار العشوائي، كما أنه التوزيع الأمثل بالصدفة والنشاط العشوائي، لذا يجب أن يكون توزيع الأداء مختلفاً تماماً عن المنحنى الاعتدالي المعياري، إذ أن الجهود التربوية تعتبر فاشلة بقدر اقتراب توزيع درجات الأفراد من المنحنى الاعتدالي. (Popham , 1978 : 84)

وقد عرض " مهرنـــز " (Mehrens, 1984) الطرق السيكومترية التقليدية في بناء الاختبارات العقلية وكيفية تحليل مفرداتها، حيث تركز هذه الطرق على فرضية أساسية وهي أن توزيع الدرجات التي يحصل عليها الأفراد على بعد القدرة التي يقيسها الاختبار للأفراد يتخذ شكل المنحنى الاعتدالي، ويمكننا تقدير النسب المئوية للأفراد في كل مستوى من مستويات هذه القدرة باستخدام دالة الكثافة الاحتمالية Probability Density Function. ولذلك يهتم باختيار المفردات التي تجعل درجات توزيع الاختبار تقترب من التوزيع الإعتدالي، وهذا يعتمد بالطبع على مدى صعوبة أو سهولة المفردات بالنسبة لعينة الأفراد، وعلى مدى تمييز كل مفردة

منها بين المستويات المختلفة للقدرة التي يقيسها الاختبار.

ولكن كثير من علماء القياس المعاصر ومن بينهم " رايت " (Wright, 1967)ولـــورد (Lord, 1968)، وهامبلتـــون (Hambleton, 1978) انتقـــدوا هذه الطـــــرق السيكومترية من جوانب أربعة هي:

١ـ جميع الخصائص السيكومترية لمفردات الاختبارات التي تستند في بنائها إلى أسس النماذج السيكومترية مثل معاملات الصعوبة، ومعاملات التمييز والثبات والصدق تختلف باختلاف طبيعة وخصائص عينة الأفراد المستخدمة، بمعنى أن معاملات الصعوبة للمفردات تتذبذب وتتغير بتغير مستوى درجات هذه السمات أو القدرات، كما تختلف باختلاف انتشار درجات هذه السمات أو القدرات، بالإضافة إلى أن الدرجات الخام المستمدة من هذه الاختبارات ليس لها معنى أو دلالة في حد ذاتها، حيث يختلف معناها ودلالتها باختلاف مدى صعوبة أو سهولة مفردات الاختبار، فإذا اشتمل اختباران على نفس العدد من المفردات التي تقيس سمة أو قدرة معينة لدى مجموعة من الأفراد، فإن درجة كل فرد ستختلف باختلاف صعوبة المفردات أو سهولتها في كل منها، وهذا يجعل الاستفادة من نتائج الاختبارات تقتصر على العينة المستخدمة التي استمدت منها هذه النتائج أو عينة مماثلة لها بقدر الإمكان.

٢ـ تقتصر الموازنة بين الأفراد في السمة أو القدرة التي يقيسها الاختبار على تطبيق نفس مفردات الاختبار أو مجموعة مفردات مكافئة لها على كل فرد من الأفراد، وبالتالي لا نستطيع الموازنة بين مستويات القدرة إذا أجاب الأفراد على مفردات مختلفة ومتباينة في صعوبتها.

٣ـ تفترض النماذج والطرق الكلاسيكية تساوي تباين أخطاء القياس لجميع الأفراد الذين يطبق عليهم الاختبار، ولكن يلاحظ أحياناً أنه بعض الأفراد يكون أداؤهم على الاختبار أكثر اتساقاً من غيرهم من الأفراد، وأن درجة هذا الاتساق تختلف باختلاف مستوى قدرة الأفراد، أو بمستوى القدرة التي يقيسها الاختبار. ولهذا

نتوقع أن أداء الأفراد ذوي القدرة المرتفعة على صورة متكافئة من صور الاختبار يكون أكثر اتساقاً من أداء الأفراد ذوي القدرة المتوسطة. (Wright, 1980 : 2) ، (علام ، ١٩٨٥: ١٠٣) ،(علام ، ١٩٨٧ : ٢٠)

٤- يمكن جمع الدرجات التي يحصل عليها فرد ما في مفردات الاختبار كما لو كانت تمثل ميزانا خطياً Linear Scale ، وأن المفردات المتعلقة بالمتغير أو الظاهرة موضع القياس تحمل نفس المعنى لدى جميع الأفراد والمفحوصين، غير أن هذا الميزان يكون منحنياً لأن الفرق الثابت بين درجتين من درجات الاختبار يختلف معناه بحسب موقع الدرجات على محور أو متصل السمة أو القدرة التي يقيسها هذا الاختبار، وهذا بالإضافة إلى أن هذه النماذج الكلاسيكية أو التقليدية تفترض أن درجات الاختبار التي تمثل السمة أو القدرة المقيسة دالة خطية مطردة. بمعنى أنه كلما زادت درجة الفرد على الاختبار دل ذلك على زيادة مقدار السمة أو القدرة لديه، ولكن ربما لا يكون هذا صحيحا في كثير من الحالات، حيث إن بعض الأفراد ذوي القدرات المرتفعة يحصلون أحياناً على درجات منخفضة في الاختبارات وقد يحدث العكس أحياناً بالنسبة للأفراد ذوي القدرات المنخفضة.(علام، ١٩٨٦ : ١٠٠).

٥ ــ لا تقدم النماذج الكلاسيكية أو الطرق السيكومترية التقليدية تفسيراً سيكولوجيا يوضح كيف يحاول الأفراد إجابة إحدى مفردات الاختبار، بالإضافة إلى أن تكوين مفردات الاختبار ومعناها تتغير بتغير عامل الزمن، أي بمضى الزمن بالنسبة لعينة الأفراد الذين أعد لهم الاختبار فإن الظروف البيئية تتغير والظروف الاختبارية ليست دائما مقننة، كما أن حذف أو تغيير مفردة من مفردات الاختبار يؤدي إلى تغيير في درجات الأفراد تغييرا يصعب التنبؤ به. (يوسف، ١٩٩١ :

٤٤٦ ـ ٤٤٧)

ولذلك فقد كانت جهود كثير من علماء القياس النفسي والتربوي في السنوات الأخيرة من أمثال " هامبلتون وآخرون " (Hambleton et al , 1978) " رايت ودوجلاس " ((Wright and Douglas, 1979) و " رايت " (Wright, 1980) ولورد (Lord , 1980) متجهة بقيادة " رايت " Wright إلى تطوير نماذج سيكومترية جديدة غير تقليدية تفي وتحقق الموضوعية في القياس سميت باسم "نماذج السمات الكامنة" Latent Trait Theory (LTT) أو نماذج الاستجابة للمفردة Item Response Theory (IRT) ولهذه النماذج مزاياها المفيدة والهامة في مواجهة الكثير من المشكلات التي عجزت عن حلها النماذج الكلاسيكية أو التقليدية، بالإضافة إلى معالجة العيوب والقصور التي لحقت بتلك النماذج والتي بيناها فيما سبق.

ويفترض في هذه النماذج الحديثة LTT أوIRT أنه يمكن التنبؤ بأداء الأفراد أو يمكن تفسير أداؤهم في اختبار نفسي أو تربوي معين بناء على خاصية أو خصائص معينة مميزة لهذا الأداء تسمى السمات Trait وتحاول هذه النظرية تقدير درجات الأفراد في هذه السمات، ونظراً لصعوبة ملاحظة هذه السمات بطريقة مباشرة أو قياسها بصورة مباشرة، فإنه يجب تقديرها أو الاستدلال عليها من خلال استجابات الأفراد الملاحظة على مفردات الاختبار الذي يقيس السمة أو القدرة المطلوبة ولهذا أطلق عليها نظرية السمات الكامنة. Latent Trait Theory (LTT).(علام، ١٩٨٢ : ٤٢)، (علام، ١٩٨٧ :٢٢)

وتعتبر نماذج السمات الكامنة نماذج أو دوال رياضية تحدد العلاقة بين أداء الفرد في مفردات اختبار وهو ما يمكن ملاحظته ملاحظة مباشرة وبين السمات أو القدرات التي يفترض أنها تكمن وراء هذا الأداء ، ولهذا تعتبر هذه الدوال احتمالية Probabilistic وليست حتمية Deterministic، وذلك يعني أن العلاقة التي تحددها الدالة تسير وفق نظرية الاحتمالات. (علام، ١٩٨٥ : ١٠٣)

والقيمة الاحتمالية لاستجابة فرد على مفردة اختبار تكون دالة لكل من السمة أو القدرة أو الخاصية Trait التي يفترض أن الاختبار يقيسها لدى الفرد وخصائص

المفردة التي يحاول الإجابة عنها، أي أنها تفترض أن هناك دالة احتمالية تربط بين بارامترين أو معلمين أحدهما يتعلق بالفرد و الآخر يتعلق بالمفردة التي يختبر بها. (علام، ١٩٩٠ : ١٢٩)

ويشير " لورد ونوفيك " (Lord and Novick , 1986) إلى أن نماذج السمات الكامنة تقوم على افتراضين أساسيين هما:

١ـ أن تكون جميع مفردات الاختبار متجانسة، أي أن جميعها تقيس سمة واحدة أو قدرة واحدة فقط.

٢ـ أن تكون إجابة الفرد عن إحدى مفردات الاختبار مستقلة إحصائياً عن إجابته على بقية المفردات في الاختبار. ولكن يجب ملاحظة أن هذا الافتراض لا يعني عدم ارتباط مفردات الاختبار بالنسبة للعينة الكلية للأفراد. (Lord and Novick, 1986 : 301)

ويعد التقدير الإحصائي للعلاقة بين احتمال الاستجابة الصحيحة على المفردة في الاختبار والسمة التي يفترض أن يقيسها هذا الاختبار هو المشكلة الرئيسية التي تواجه من يستخدم هذه النظرية في بناء الاختبارات والمقاييس التربوية والنفسية، وتمثل هذه العلاقة كل مفردة من مفردات الاختبار بمنحنى يطلق عليه المنحنى اللوغاريتمي التراكمي logistic curve (Lord and Novick , 1986) كما تتطلب الحصول على معلومات من مصدرين أحدهما يتعلق بالفرد المختبر، والآخر يتعلق بمفردة الاختبار، وعادة يتطلب الأمر قيمة عددي واحدة تتعلق بالفرد (معلم القدرة أو السمة أو الخاصية المقاسة لدى الفرد، وقيمة عددية أو أكثر تتعلق بالمفردة (معلم أو معالم المفردة) واختلاف عدد معالم مفردات الاختبارت يرجع إلى اختلاف منظور علماء القياس فيما يتعلق بنمذجة العملية الاختبارية، فمن هؤلاء العلماء من يرى أنه ينبغي أن تكون للمفردة معلم واحد هو صعوبة المفردة Item Diffuculty Parameters، ومنهم من يضيف إلى ذلك معلم آخر هو تمييز المفردة Item Discrimination Parameter وآخرون يرون أنه من الأفضل إضافة معلم ثالث يأخذ في الاعتبار عامل تخمين الاستجابة وخاصة في مفردات الاختيار من متعدد Guessing Parameter لذلك تعددت النماذج المتعلقة بهذه النظرية وتباينت أشكال

المنحنيات المميزة تبعاً لإختلاف عدد معالم المفردات، فنموذج راش Rasch Model هو النموذج أحادي البارامتر أو المعلم، ويفترض هذا النموذج توازي المنحنيات المميزة لجميع مفردات الاختبار، ونموذج لورد Lord Model هو نموذج ثنائي المعلم، ونموذج بيرنيوم Birnbum Model هو نموذج ثلاثي المعلم، وهذان النموذجان يسمحان للمنحنيات المميزة للمفردات بالتقاطع. (علام، ١٩٩٠ : ١٣١).

ولكل من هذه النماذج مميزات وعيوب، غير أن نموذج راش يعد أبسط هذه النماذج وأكثرها استخداماً في التطبيقات العلمية وبخاصة في مجال بناء الاختبارات العقلية والتحصيلية والاستبانات كما يعد أكثر هذه النماذج التي أجريت حولها دراسات أجنبية متعددة للتحقق أمبريقياً من خصائصه وجدواه في مختلف المجالات السيكومترية التي يصعب معالجتها استناداً إلى أسس ومبادئ النظرية الكلاسيكية. (علام، ١٩٩٠ : ١٣١)

ويتميز نموذج راش كذلك عن غيره من نماذج السمات الكامنة بثلاث مميزات أساسية تتعلق بمفهوم الموضوعية Objectivity في القياس والذي نادى به رايت (Wright, 1967, 1975, 1976, 1977, 1979) وهذه المميزات هي:

١ـ إذا افترضنا توفر عينة كبيرة من مفردات الاختبار بحيث تقيس جميعها نفس القدرة أو السمة، فإنه يمكن الحصول على قيم تقديرية Estimate لقدرة الفرد، وتكون هذه القيمة مستقلة احصائياً عن عينة المفردات التي اختبر بها الفرد. (Item – free person measurement).

٢ـ إذا افترضنا توفر عينة كبيرة من الأفراد فإنه يمكن الحصول على قيم تقديرية Estimate لمعاملات الصعوبة، ومعاملات التمييز للمفردات التي اختبروا بها وتكون هذه القيم مستقلة إحصائياً عن عينة الأفراد التي أجرى عليها الاختبار (Person Free test calibration).

٣ـ يمكن استخدام النموذج للحصول على معامل إحصائي يدل على مدى دقة تقدير قدرة كل فرد، وربما تختلف قيمة هذا المعامل من فرد إلى أخر بحسب مستوى قدرة كل منهم. (علام، ١٩٨٦ : ١٠٣)

وعلى الرغم من أن نموذج راش يعتبر حالة خاصة من النموذج ثنائي المعلم إلا أنه يتميز بخصائص تجعله أكثر تطبيقاً واستخداماً في تحليل مفردات الاختبارات

عن غيره من نماذج السمات الكامنة، ومن هذه الخصائص:

١ـ يعتبر أقل النماذج في عدد الافتراضات اللازم تحققها في البيانات المستمدة من الاختبار لكي نحصل باستخدامه على تقديرات Estimates دقيقة لكل من صعوبة المفردة Difficulty وقدرة الفرد Person Ability.

٢ـ تمكن علماء القياس من إيجاد حلول إحصائية مناسبة لمشكلة تقدير معالم النموذج ModelParameters في حين أنهم يواجهون حتى الآن مشكلات سيكومترية وإحصائية في تقدير معاملات النماذج الأخرى.

٣ـ تتساوى القيمة التقديريرة لقدرات الأفراد Ability Estimates الذين حصلوا على نفس الدرجة الكلية في الاختبار بغض النظر عن عينة المفردات التي اختبر بها كل منهم، في حين أنه ليس من الضروري حدوث ذلك في حالة النماذج الأخرى. (علام، ١٩٨٥ : ١٠٥ ـ ١٠٦)

وبناء على كل ما سبق، فقد حاز نموذج "راش" على اهتمام معظم الباحثين وعلماء القياس المعاصر وبخاصة في مجال بناء الاختبارات والمقاييس النفسية والتربوية بأنواعها المختلفة في مجالات عديدة أخرى كبنوك الأسئلة على سبيل المثال. وأصبح هذا النموذج يسود ويسيطر على الفكر النفسي والتربوي في مجال القياس تمشياً مع الحركة التطورية العالمية السريعة التي انتشرت في كثير من دول العالم ككندا و أستراليا والسويد، وغيرها من الدول الأخرى التي جعلت من هذا النموذج واقع فعلي في كافة مجالات القياس المختلفة وبخاصة في مجال العمليات التعليمية والتربوية. (يوسف، ١٩٩١: ٤٥٤)

ونظراً لهذه المميزات التي يتميز بها نموذج راش عن غيره من نماذج السمات الكامنة الأخرى، فضلاً عن بساطة وسهولة تقدير بارامتريه باستخدام الآلة الحاسبة دون الحاجة إلى استخدام الحاسب الآلي، لذلك سوف تعتمد هذه الدراسة عليه في تحليل بياناتها.

نموذج راش أحادي البارامتر أو المعلم: Rasch Model (one Parameter Model)

يعتبر هذا النموذج حالة خاصة من نموذج بيرنبوم ثنائي البارامتر فهو يقوم ـ

نموذج راش - على افتراض أن جميع مفردات الاختبار تميز بنفس القدر بين الأفراد ولكنها تختلف عن بعضها في الصعوبة فقط. ويهدف هذا النموذج إلى تحقيق الموضوعية في القياس Objectivity بمعنى أن درجة الفرد في الاختبار لا يجب أن تكون دالة لعينة الأفراد الذين استخدموا في التدريج الأصلي للمفردات Item Calibration التي يشتمل عليها الاختبار كما أنه يجب أن يحصل الفرد على نفس الدرجة في كل من اختبارين يقيسان نفس السمة أو القدرة. ولتوضيح ذلك نفترض أننا قمنا ببناء وتطبيق اختبارين يشتملان على نفس العدد من المفردات التي تقيس نطاقا سلوكياً معيناً بناء على النظرية الكلاسيكية في القياس على مجموعة معينة من الأفراد وبفرض أن أحد الاختبارين يشتمل على المفردات الصعبة، والآخر يشتمل على المفردات السهلة فسوف نجد أن درجة فرد معين من أفراد هذه المجموعة في كل من الاختبارين يختلف باختلاف صعوبة المفردات حيث يتوقع أن يحصل الفرد على درجة أعلى في الاختبار الأقل صعوبة، بينما يحصل على درجة أقل في الاختبار الأكثر صعوبة، وهذا يعني أن القياسات الناتجة عن مثل هذه الاختبارات تعتمد على المفردات التي يشتمل عليها كل اختبار منهما، أي أن هذه القياسات غير متحررة أو مستقلة عن المفردات المستخدمة في القياس Not Item Free وكذلك إذا استخدمنا في تقنين اختبار معين عينة من تلاميذ المرحلة الابتدائية مثلاً، فإنه لا يجوز تطبيق هذا الاختبار (بعد تقنينه) على تلاميذ مرحلة مختلفة كالثانوية العامة وذلك طبقاً للنظرية السيكومترية الكلاسيكية، إذ يجب أن تكون خصائص الأفراد في عينة التقنين مماثلة لخصائص أفراد المجتمع المستهدف، وبذلك لا تكون أدوات القياس التي تبنى على هذا الأساس مستقلة عن خصائص العينة Not Sample Free (يوسف، ١٩٩١: ٤٤٩)

ولذلك يذكر "راش" أن هذا المفهوم الموضوعي للقياس غير متوافر في

~~القياس النفسي والتربوي كما هو متوافر في العلوم الطبيعية. ولكي يتحقق شرط~~ الموضوعية في القياس النفسي والتربوي يجب أن تتحرر تداريج أدوات القياس من خصائص الأشياء المراد قياسها وأن يتحرر قياس الأشياء من خصائص أدوات القياس، حيث إن توافر هذان الشرطان يمكننا من تعميم القياس، بحيث لا يقتصر على أداة قياس بعينها، كما يمكننا من موازنة الأشياء المراد قياسها باستخدام أدوات قياس متشابهة وليست بالضرورة متطابقة. (Willmot and Fowles, 1974: 15)

معنى الموضوعية في نموذج راش:

تعني الموضوعية في نموذج راش، موضوعية المقارنة بين نتيجة تفاعل قدرتي فردين على مفردة مناسبة، أي موضوعية المقارنة بين استجابتي فردين على مفردة مناسبة، كما تعني أيضاً موضوعية المقارنة بين صعوبة مفردتين أجاب عليهما فرد مناسب، وتتمثل متطلبات هذه الموضوعية في جانبيين هما:

١ـ استقلال معلم قدرة الفرد عن البند المستخدم Item Free .

٢ـ استقلال معلم صعوبة البند عن الفرد الذي يجيب عليها Person Free .

(الشرقاوي وآخرون، ١٩٩٦ : ٣٢١ ـ ٣٢٢)

و الصورة الرياضية الممثلة لاحتمال الإجابة الصحيحة عن إحدى المفردات في الاختبار لنموذج راش، تأخذ الصورة :

$$ل = \frac{١}{١ + e^{ص_م - ق}} \quad، م = ١ ، ٢ ، ٣ ، .... ن.$$

حيث (ل) ترمز إلى احتمال أن يجيب فرد إجابة صحيحة عن مفردة (م) ، (ص م) ترمز إلى صعوبة المفردة (م)، (ق) ترمز إلى مقدار السمة الكامنة (القدرة الحقيقية) كما تقاس بالاختبار. ، e ترمز إلى الأساس اللوغاريتمي الطبيعي. (علام، ١٩٩٥ : ١٨٦)

وعلى ذلك فإنه يمكن تعريف قدرة الفرد بأنها اللوغاريتم الطبيعي لترجيح الإجابات الصحيحة للفرد على البنود التي تعبر نقطة صفر التدريج عن صعوبتها.

(الشرقاوي وآخرون، ١٩٩٦: ٤٤٥).

تعقيب عام :

يتضح من خلال العرض السابق لمفهوم الدرجة الحقيقية، ومفهوم الثبات وعلاقته بالدرجة الحقيقية، ونظرية السمات الكامنة (نموذج راش) أن هناك اختلاف في كل من المدخلين، المدخل الكلاسيكي التقليدي والذي يعتمد على معاملات الثبات ومتوسط الدرجات الخام في تقدير الدرجات الحقيقية، والنماذج الحديثة وبصفة خاصة نموذج راش والذي يهدف إلى تحقيق الموضوعية في القياس من خلال تقدير السمات الكامنة (الدرجات الحقيقية) للأفراد، والتي تقف خلف إجاباتهم عن مفردات الاختبار المختلفة. فهل فعلاً يتم قياس الحقيقة لقدرات الفرد أو السمات الكامنة لهم في إجاباتهم عن الاختبارات والمقاييس المختلفة؟ وهل يختلف تقدير هذه الدرجات الحقيقية باختلاف المدخلين الكلاسيكي التقليدي والمدخل الحديث أو نظرية السمات الكامنة؟ هذا ما تحاول هذه الدراسة الكشف عنه.

الدراسات السابقة:

نظراً لأن الباحثة لم تجد دراسات سابقة تناولت موضوع الدراسة بصورة مباشرة، فقد تم تقسيم الدراسات السابقة التي تم الحصول عليها في المحورين التاليين:

المحور الأول: دراسات اهتمت بنماذج السمات الكامنة (نموذج راش)

المحور الثاني: دراسات اهتمت بالثبات.

المحور الأول: دراسات اهتمت بنماذج السمات الكامنة (نموذج راش):

١- دراسة علام (١٩٨٢):

حيث تناولت الدراسة أحد أهم التطورات المعاصرة في القياس والتقويم النفسي والتربوي والذي يرتبط مباشرة ببناء وتنظيم بنوك الأسئلة وذلك من خلال

التعريف ببنوك الأسئلة والتفرقة بين نوعين منها، بالإضافة على تناول لأهم أوجه النقد الموجه للنظرية التقليدية في القياس ثم التعريف بنماذج السمات الكامنة ومميزاتها، ثم عرض لنموذج راش الاحتمالي البسيط Rasch Simple Logistic Model وأسسه الرياضية وكيفية الاستفادة من هذا النموذج في بناء بنوك الأسئلة.

٢- دراسة رونالد وداتو (Ronald and Dato, 1983)

هدفت الدراسة إلى مقارنة نتائج انتقاء مفردات اختبار مرجعي المحك باستخدام نماذج السمات الكامنة (أحادي البارامتر - ثنائي البارامتر) وذلك عن طرق المحاكاة بالكمبيوتر، وبينت نتائج الدراسة أفضلية نموذج بيرنيوم في الانتقاء للمفردات.

٣- دراسة "رونالد " وآخرين (Ronald and others, 1983)

هدفت الدراسة إلى الاستفادة بإجراءات المحاكاة بالكمبيوتر ونظرية الاستجابة للمفردة في انتقاء مفردات الاختبار مرجعي المحك وتحديد طوله وانتقاء درجة القطع المناسبة والمقارنة بين طول الاختبار ومؤشرات ثبات وصدق الاختبار، وبينت نتائج الدراسة دقة نماذج السمات الكامنة (راش) في انتقاء المفردات المناسبة.

٤ - دراسة علام (١٩٨٥):

هدفت الدراسة إلى تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي، وقد طبق اختبار للذكاء العام (أحد اختبارات بطارية الاستعدادات العقلية)على عينه مكونة من ١٢٦ طالباً من بين الملتحقين الجدد بالسنة الأولى في مختلف أقسام كلية التربية - جامعة الأزهر بالقاهرة، واستخدام الباحث طريقة " بروكس " PROX للتحقق من فروض دراسته وهي: ١- تنطبق البيانات المستمدة من هذا الاختبار على نموذج راش. ٢ـ لا يتغير تدريج الاختبار بتغير

مستوى قدرات عينة الطلاب المستخدمة للحصول على هذا التدريج. وقد بينت نتائج التحليل أن القيم التقديرية لكل من صعوبة المفردات وقدرة الطلاب تطابق النموذج. كذلك بينت نتائج الدراسة أن مفردات الاختبار مستقلة إحصائياً عن قدرة عينة الطلاب المستخدمة في الحصول على هذه القيم، وبهذا أكدت النتائج أهمية استخدام راش في القياس الموضوعي لقدرات الطلاب.

٥ ـ دراسة كاظم (١٩٨٨ أ)

هدفت الدراسة إلى تقديم دراسة نقدية مفصلة حول القياس الموضوعي من خلال توضيح مفصل لأحد الاتجاهات الجديدة في القياس الموضوعي للسلوك الذي يختص بأهم نماذج السمات الكامنة وهو نموذج راش، وتوضيح كيفية تحقيق متطلبات الموضوعية في تفسير نتائج القياس بناء على نموذج راش، ومناقشته لأهم التطبيقات العملية لنموذج (راش) في مجال القياس السلوكي. وكيف يمكن الاستفادة من ذلك في حل مشاكل القياس في بيئاتنا العربية، وأخيراً تقديم مناقشة نقدية حول استخدام نموذج راش في تفسير نتائج القياس.

٦ ـ دراسة كاظم (١٩٨٨ ب):

هدفت الدراسة إلى استخدام نموذج راش في بناء اختبار موضوعي في علم النفس، وتفسير أداء الأفراد على هذا الاختبار تفسيراً موضوعياً، ثم التحقق من مدى موضوعية هذا التفسير، كما يتمثل ذلك في تحقيق فرضيات نموذج (راش)، وتناولت الباحثة في هذه الدراسة تلخيصاً لمطالب الموضوعية في القياس، وتلخيصاً لنموذج راش يتضمن فروض النموذج والصيغة الرياضية له، ومعنى الموضوعية في نموذج (راش) وكيفية تقدير كل من معلم صعوبة المفردات وقدرات الأفراد بطريقتين مختلفين هما طريقة PROX وطريقة UCON -تقدير الترجيح الأعظم غير

المشروط - وتم إجراء التحليلات المناسبة للبيانات باستخدام الطريقتين السابقتين وخلص هذا التحليل إلى مناسبة بنود المقياس المعد في مادة علم النفس لنموذج راش مع استبعاد بعض العبارات غير المناسبة أو التي لا تنطبق عليها شروط النموذج.

٧ ـ دراسة كاظم (١٩٨٨ ج):

هدفت الدراسة إلى معرفة أثر(تقارب ـ تباعد) مستوى عينات التحليل مع مستوى كل من اختبارين أحدهما سهل والآخر صعب على دقة عملية التعادل الرأسي لتدريجها، ومن ثم تدريجها في تدريج واحد مشترك وصفر واحد مشترك. واستخدمت الباحثة اختباراً موضوعياً في علم النفس تم تدريجه بطريقة نموذج (راش)، وتكون الاختبار في صورته النهائية بعد حذف البنود غير الملائمة من (٩٥) بندا صالحة للقياس، ثم تقسيم الاختبار الكلي إلى اختبارين فرعيين يضم كل منهما (٥٥) بنداً ويشتركا في (١٤) بنداً تستخدم كبنود رابطة بينهما وذلك على الصورة الآتية: بالنسبة للاختبار السهل تكون من (٤٠) بنداً السهلة بالإضافة إلى (١٤) بندا الرابطة، أما الاختبار الصعب فيتضمن الـ (٤٠) بنداً الصعبة بالإضافة إلى (١٤) بنداً الرابطة، وتكونت عينة الدراسة من (٤١٨) طالباً وطالبة يدرسون مقرر مدخل في علم النفس بجامعة الكويت، وتم تقسيم هذه العينة إلى عينة مرتفعة المستوى وعددهم (٢٠٥) وعينة منخفضة المستوى (٢١٣) و تشترك العينتان في (١٧) فرداً. واستخدم برنامج بيكال BICAL في تحليل البيانات بطريقة نموذج راش. وبينت نتائج الدراسة إلى أهمية تقارب مستوى الاختبار مع مستوى عينة التحليل لكي تصل بذلك إلى الدقة والتقديرات التي تؤدي إلى الدقة في عملية التعادل الرأسية في تدريج الاختبارات.

٨ ـ دراسة علام (١٩٩٠):

هدفت الدراسة إلى استخدام نموذج راش في بناء وتصميم مقياس هدفي المرجع لقياس مدى تمكن طلاب الدراسات العليا بكلية التربية بجامعة الأزهر من المعارف الأساسية المتعلقة بإعداد خطة البحوث النفسية التربوية وتصميم ميزان لوغاريتمي خطى يمكن في ضوئه تحديد مستويات التمكن المتوقعة من الطالب في كل هدف من الأهداف السلوكية التي تقيسها المفردات الخاصة بالمقياس. وقد حدد الباحث هذه الأهداف في ضوء التحليل السلوكي للمكونات التي ينبغي توافرها في خطة البحوث النفسية التربوية، وقام الباحث بكتابة مفردات اختبارية تقيس هذه الأهداف، كما تحقق من الصدق الوصفي للمقياس اعتماداً على أراء المحكمين،ثم أجرى دراسة تجريبية على عينة مكونة من (٤٨) طالباً من طلاب الدراسات العليا بكلية التربية جامعة الأزهر. واستخدام طريقة PROX في تحليل مصفوفة البيانات الاختبارية المستمدة من هذه العينة للحصول على القيم التقديرية لكل من صعوبة المفردات وقياسات الطلاب على متصل متغير المعارف الأساسية مقدرة بوحدة الترجيح اللوغاريتمي (اللوجيت). وصمم الباحث ميزاناً متدرجاً لمستويات التمكن المتوقعة لبعض الأهداف السلوكية للمقياس مستعيناً في ذلك، بالقيم التقديرية لكل من صعوبة المفردات وقياسات الطلاب التي تم الحصول عليها، وتطبيق الصيغة الرياضية لنموذج راش، وناقش الباحث نتائج الدراسة في ضوء هذا الميزان.

٩ ـ دراسة يوسف (١٩٩١):

حيث هدفت الدراسة إلى استخدام نموذج راش اللوغاريتمي الاحتمالي في تحليل مفردات الاختبارات المعرفية مرجعية المعيار ثنائية القطب (اختبار الأشكال المتضمنة) الصورة الجمعية من إعداد أنور الشرقاوي، وسليمان الخضري الشيخ

(١٩٨٨)، حيث طبق الباحث اختبار الأشكال المتضمنة على عينة من طلاب وطالبات الفرقتين الثالثة والرابعة بكلية التربية، جامعة المنيا والذين بلغ عددهم (٣٢٨) طالباً وطالبة، وبعد التأكد من صدق وثبات المقياس قام الباحث بتحليل مفرداته في ضوء نموذج راش باستخدام طريقة PROX. وخلصت الدراسة إلى أن البيانات المشتقة من اختبار الأشكال المتضمنة (الصورة الجمعية) والمعد من أجل قياس بعدي الاعتماد الاستقلال عن المجال الإدراكي كأسلوب معرفي ثنائي القطب، تتفق مع نموذج راش، وأن القيم التقديرية لصعوبة مفرداته لا تتغير بتغير مستوى قدرة الأفراد المستخدمة في الحصول على هذا التدريج، كما أثبتت نتائج الدراسة أيضاً أنه يمكن اختبار الطالب بأي عدد من تلك المفردات المثالية للحصول على تقدير دقيق لقدراته، كما يمكن اختيار الطلاب بعينات مختلفة من هذه المفردات الملائمة، هذا فضلاً عن أنه يمكن الموازنة بين قدراتهم بميزان لوغاريتمي مشترك.

١٠ ـ دراسة مراد (١٩٩٨):

هدفت الدراسة إلى الإجابة عن التساؤلات الآتية:

١ـ هل تتأثر دقة وكفاءة ضم مفردات اختبارين (سهل ، صعب) باختلاف حجم عينتي التحليل لكل من الاختبار السهل والاختبار الصعب (باستخدام طريقة المفردات المشتركة)؟

٢ـ هل تتأثر دقة وكفاءة ضم مفردات اختبارين (سهل ، صعب) باختلاف حجم عينة التحليل عند استخدام طريقة المفردات المشتركة إذا كانت عينتي التحليل (مختلفتي الحجم، متساويتين)، وقام الباحث باستخدام اختبار في مادة الجبر مكون من (٧٤) مفردة واشتق من هذا الاختبار اختبارين فرعيين أحدهما سهل (٣٠ مفردة) والآخر صعب (٣٠ مفردة) واستخدم الباحث عينة قوامها (١٣٠٠) طالب قسمت على النحو التالي (٨٠٠ طالب للاختبار السهل، ٤٠٠ للاختبار الصعب)، (٤٠٠

طالب للاختبار السهل، ٨٠٠ طالب للاختبار الصعب) (عينتان متكافئتان حجم كل منهما ١٠٠ ، ٢٥٠ ، ٤٠٠ طالب) واستخدام الباحث تحليلات نموذج راش باستخدام برنامج "ميكروسكيل" وقد أسفرت نتائج الدراسة عن: عدم وجود فروق ذات دلالة إحصائية بين استخدام عينتين (٨٠٠ فرد للاختبار السهل و ٤٠٠ للاختبار الصعب) وبين استخدام عينتين (٤٠٠ فرد للاختبار السهل، ٨٠٠ للاختبار الصعب) مما يدل على عدم تأثير حجم العينة على دقة إجراء الضم، كما أسفرت نتائج الدراسة عن وجود فروق طفيفة بين استخدام عينتين حجم كل منها (٨٠٠ للاختبار الصعب، ٤٠٠ للاختبار السهل) وبين عينتين متساويتين حجم كل منها (٤٠٠) لكل من الاختبار السهل والاختبار الصعب وذلك لصالح الإجراء الأول (عينتين مختلفتين في الحجم) مما يشير إلى أفضلية استخدام هذا الإجراء مع اختلاف حجم عينتي التحليل، كما أسفرت النتائج عن عدم دقة إجراء الضم (عند استخدام طريقة المجموعة المكافئة) مع اختلاف حجم عينتي التحليل (١٠٠ ، ٢٥٠، ٤٠٠ طالب) حيث أدت جميعها إلى نتائج غير دقيقة.

١١ ـ دراسة "هاندس ولاركين" (Hands and Larkin, 2001)

هدفت الدراسة إلى التحقق من مدى كفاءة بنود مقياس القدرة الحركية العامة باستخدام نموذج راش. وعلى عينة مكونة من (٣٣٢٥) في سن السادسة من الأطفال بأستراليا قاموا بأداء ٢٤ مهارة حركية تم تصنيف البيانات الخاصة بهم، وباستخدام التحليلات الخاصة بنموذج راش أمكن التوصل إلى مقياس ذا كفاءة سيكومترية مرتفعة كما تم التوصل إلى مقياسين كل منهما أحادي البعد، أحد هذين المقياسين للطلاب والآخر للطالبات.

١٢ ـ دراسة ووف (Waugh, 2001):

هدفت الدراسة إلى استخدام نموذج "راش" في تنقيح استبيان طلاب الجامعة المرتبط بخبراتهم Experiences والذي تم تطبيقه على عينة من طلاب الفرقة الأولى بالجامعة في أستراليا. وفي ضوء نموذج "راش" تم استبعاد (٥٨) عبارة من عبارات الاستبيان وأصبح الاستبيان في صورته النهائية مكوناً من (٦٦) عبارة ذات

خصائص سيكومترية ممتازة.

١٣ ـ دراسة جوزيف آخرون (Joseph and others, 2001):

هدفت الدراسة إلى مقارنــة طريقة ليكرت، وطريقة راش في تدريــج بنـود المقاييس المختلفة. وتـم تطبيق مقياس أنشطة الحيـاة اليوميـة على عينة مكونـة من (٤٤٣٠) من الشباب و (٦٠٥) من ذوي الإعاقة البدنية. وبتحليل بيانات العيتنين في ضوء تقدير ليكرت وطريقة راش بينت النتائج تفوق نموذج راش على طريقة ليكرت في انتقاء المفردات ذات الكفاءة السيكومترية.

١٤ ـ دراسة جارنر وانجليهارد (Garner and Engelhard, 2002) :

هدفت الدراسة إلى قياس مدى كفاءة طريقة ايجنفاكتور Eiegenvector في تقدير بارامتري نموذج راش" صعوبة المفردة و قدرة الفرد" للبيانات الثنائية التي تعتمد علـى (٠ ، ١) والبيانـات غير الثنائيـة Polytomous وبينت نتـائج الدراسـة إمكانية استخدام هذه الطريقة في تقدير بارامتري نموذج راش سواء كانت البيانات ثنائية (Dicotomous) أو غير ثنائية Polytomous.

تعقيب على دراسات المحور الأول:

على الرغم من تعدد الدراسات التي تناولت نموذج راش كأحد أهم نماذج السمات الكامنة، واختلاف أهداف هذه الدراسات ما بين استخدام النموذج في بناء بنوك الأسئلة كما في دراسة علام (١٩٨٢)، أو استخدام النموذج في تحليل بيانات الاختبارات العقلية والمعرفية كما في دراسة علام (١٩٨٥) و يوسف (١٩٩١) أو استخدام النموذج في بناء الاختبارات التحصيلية كما في دراسة علام (١٩٩٠) و كـــــــــــــــاظم (١٩٨٨ ب)، ودراسة رونالـــد وداتـــــو (Ronald and Dato, 1983) ودراسة هاندس ولاركن (Hands and Larkin , 2001). وغير هذه الدراسات والتي أثبتت

كفـاءة نمـوج راش فـي بنـاء وانتقـاء مفـردات المقـاييس النفـسية والتربويـة ذات الخصائص السيكومترية الممتازة. إلا أنه لم توجد دراسة واحدة تناولت العلاقة بين القدرة الحقيقية كما تقدر بنموذج راش والدرجات الحقيقية المقدرة باستخدام النظرية الكلاسيكية في القياس. كذلك يلاحظ من خلال هذه الدراسات أن طريقة بروكس PROX هي أكثر الطرق استخداما في معالجة البيانات باستخدام نموذج راش.

المحور الثاني: دراسات اهتمت بالثبات:

على الرغم من كثرة وتعدد الدراسات السابقة التي تناولت موضوع الثبات إلا أنها تناولته كأحد الخصائص السيكومترية للمقاييس النفسية وذلك بالإضافة إلى الصدق، كما في دراسة شارفي (Scharfe, 2002) ودراسة توماس ورونالد (Thomas & Ronald , 2002) ودراسة كارسو وتكوتيز (Caruso & Witkiewitz, 2002) وغير هذه الدراسات الكثير. إلا أن الباحثة استطاعت الحصول على بعض الدراسات السابقة والتي اهتمت ببعض الموضوعات المتعلقة بالثبات ومن أمثلة هذه الدراسات:

١- دراسة رويس واليوت (Royce & Eliot, 1978)

هدفت الدراسة إلى بحث العلاقة ببين تضخم الدرجات وقيم معاملات الثبات للاختبار، وتوصلت الدراسة إلى أن الزيادة في متوسط الدرجات ليس بالضرورة أن يتبعه أو يلازمه انخفاض في ثبات هذه الدرجات، أو ضعف التمييز في القدرة الحقيقية للطلاب ذوي المستويات المختلفة أو انخفاض في القوة التنبؤية لهذه الدرجات.

٢ ـ دراسة كوميتا وويليم (Komorita & William, 1985)

هدفت الدراسة إلى بحث العلاقة بين عدد درجات الاختبار Numbre of scale points والثبات وبينت نتائج الدراسة عدم وجود علاقة بين عدد درجات الاختبار وثباته إذا كانت بنود هذا الاختبار متجانسة Homogeneous ولكن

في الاختبارات غير المتجانسة Heterogeneous البنود يزيد ثبات هذه الاختبارات ليس فقط بزيادة عدد بنوده ولكن أيضاً بزيادة عدد درجات تلك البنود أو تقديراتها.

٣ ـ دراسة الصياد (١٩٨٨):

هدفت هذه الدراسة إلى محاولة العثور على مؤشر إحصائي يمكن استخدامه في تقدير الثبات بطريقة تطبيق وإعادة التطبيق لأداة القياس، بغض النظر عن مستوى القياس Level of Measurement الذي يميز القياسات التي تمدنا بها الأداة المطلوب قياس ثباتها، وكذلك هدفت الدراسة إلى مقارنة هذا المؤشر الجديد وبين المؤشرات الأخرى الشائعة الاستخدام وتلك الممكنة الاستخدام كمؤشرات لتقدير الثبات بطريقه التطبيق وإعادة التطبيق، وتناول الباحث في هذه المقارنة المميزات والعيوب لكل المؤشرات الداخلة في المقارنة.

٤ـ دراسة الزيات (١٩٨٩):

هدفت الدراسة إلى الكشف عن أثر قوة تثبيت البدائل في أسئلة الاختيار من متعدد على معاملات الصعوبة، معاملات التمييز، الانحراف المعياري للفقرات وللاختبار ككل، الصدق العاملي والصدق الفارقي للاختبار وثباته، وخلصت الدراسة إلى أن لقوة مشتتات الإجابة تأثير دال على معاملي الصعوبة والتمييز في أسئلة الاختيار من متعدد، كما يمكن زيادة ثبات الاختبارات القائمة على الاختيار من متعدد بزيادة فاعلية مشتتات الإجابة، كما يمكن زيادة الصدق الفارقي والصدق الإرتباطي للاختبارات القائمة على الاختيار من متعدد وبزيادة فعالية مشتتات الإجابة.

٥ ـ دراسة الصياد (١٩٩٠):

هدفت الدراسة إلى بحث العلاقة بين قيمة المعامل ألفا لكرونباخ ونوعية استجابة المستجيب لعبارة "اكتب اسمك إذا رغبت" ومن خلال الحصول على البيانات الخام لثلاثة دراسات أجراها المؤلف مع آخرين أمكن الحصول على

مجموعتين للثبات بطريقة المعامل ألفا لكرونباخ أحدهما حسبت من خلال الذين عرفوا أسماءهم والثانية من الذين لم يعرفوا أسماءهم. وبينت نتائج الدراسة أنه لا توجد فروق ذات دلالة إحصائية بين متوسطات ثبات المجموعتين. كما أنه لا توجد فروق ذات دلالة إحصائية وعملياً في نفس الوقت بين متوسط استجابات المجموعتين.

٦ ـ دراسة غنيم (١٩٩١):

هدفت الدراسة إلى تحديد معادلة عامة تصلح لحساب معامل ثبات أي مقياس سواء كانت مفرداته متجانسة أو غير متجانسة؟ وفي حالة التوصل لمثل هذه المعادلة فهل يمكن التحقق من أن معامل ألفا حالة خاصة منها؟ وما إذا كانت هناك اختلافات بين معامل الثبات المحسوب بالمعادلة العامة عن نظيره المحسوب بمعامل ألفا باختلاف درجة تجانس المفردات؟ وأخذت البيانات من تطبيق خمسة مقاييس نفسية (الذكاء المصور لأحمد زكي صالح، الخجل لعبد العزيز الدريني، والاكتئاب، لغريب عبد الفتاح، اختبار الشخصية النرجسية لعبد العزيز أحمد البحيري، واختبار القدرة العددية من إعداد خليل ميخائيل معوض) على عينة مكونة من (٤٧) طالباً من طلاب جامعة الإمام محمد سعود الإسلامية بالقصيم بالمملكة العربية السعودية. وأمكن للباحث التوصل إلى معادلة عامة لحساب ثبات المفردات المتجانسة وغير المتجانسة وكان معامل ألفا حالة خاصة منها بالإضافة إلى اختلاف قيم معاملات الثبات المقدرة بالمعادلة العامة عن نظيرها المقدرة بطريقة ألفا اختلافا طفيفا.

٧ ـ دراسة غنيم (١٩٩٢):

هدفت الدراسة إلى محاولة التوصل إلى صيغ تقريبية لمعامل ألفا لتقدير معامل ثبات المقاييس ذات المفردات المتجانسة تسهيلاً للباحثين في تقدير الثبات

بطريقة ألفا وأمكن للباحث من التوصل إلى ثلاث صيغ تقريبية لمعامل ألفا يتم اختيار إحـداها وفقـاً لبيانـات المقيـاس (النهايـة العظمـي لـدرجات المقيـاس، عـدد مفـردات المقياس، بالإضافة إلى الانحراف المعياري لدرجات المقيـــاس ووزن كل مفـــردة) وقد عمـــل الباحث على تحسين الصيغ التقريبية لمعامل ألفا مستفيداً من نهج هورست Horst (١٩٦٦) واستانلي stanly في تحسين معادلة كيودر ريتشاردسون رقم (٢١) ودلت النتائج النهائية على وجود فروق طفيفة بين معامل الثبات المحسوب بالصيغ التقريبية المحسنة ونظيره المحسوب بمعامل ألفا وهذه الفروق لا تتعدى ٠,٠٢.

تعقيب على دراسات المحور الثاني:

على الرغم من اختلاف أهداف الدراسات السابقة التي تناولت مفهوم الثبات إلا أنها ركزت على طرق تقدير الثبات ومحاولة التوصل إلى طرق جديدة لحساب الثبات لتلاقي العيوب في بعض الطرق القديمة، أو أن الواقع التربوي للبحوث دعت إلى الحاجة إلى طرق جديدة وذلك كما في دراسة الصياد (١٩٨٨)، الصياد (١٩٩٠) وغنيم (١٩٩١) وغنيم (١٩٩٢). والبعض الآخر من الدراسات ركز على العلاقة بين الثبات وبعض المتغيرات الأخرى كما في دراسة الزيات ١٩٨٩ التي ركزت على أثر قوة تشتت البدائل على الثبات، ودراسة رويس واليوت (Royce and Eliot, 1978)التـي بحثـت العلاقـة بـين تـضخم الـدرجات وقـيم معـاملات الثبـات ودراسـة (Komorita & William, 1985).التي بحثت العلاقة بين عدد درجات الاختبار وقيم معامل القياس . وعلى الرغم من أهمية هذه الدراسات والحاجة إليها إلا أنه لا توجد دراسة اهتمت ببحث العلاقة بين الدرجات الحقيقية المقدرة من نموذج راش، وتلك المقدرة بالنماذج التقليدية أو الكلاسيكية.

فروض الدراسة:

الفرض الأول: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للطلاب والطالبات على مقياس اليقظة العقلية.

الفرض الثاني: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للتخصصات الأدبية والتخصصات العلمية على مقياس اليقظة العقلية.

الفرض الثالث: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (المقدرة بنموذج راش) للطلاب والطالبات على مقياس اليقظة العقلية.

الفرض الرابع: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (المقدرة بنموذج راش) للتخصصات العلمية والأدبية على مقياس اليقظة العقلية.

الفرض الخامس: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية للطلاب والطالبات على مقياس اليقظة العقلية.

الفرض السادس: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية للتخصصات العلمية والتخصصات الأدبية في مقياس اليقظة العقلية.

الفرض السابع: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بنموذج راش وتلك المقدرة بالمدخل الكلاسيكي لمقياس اليقظة العقلية.

إجراءات الدراسة:

١ ـ منهج الدراسة:

استخدم المنهج الوصفي الإحصائي لملائمته لطبيعة وبيانات الدراسة.

٢- مجتمع الدراسة:

يتكون مجتمع البحث من جميع الطلبة المسجلين بكلية التربية في جامعة الكويت في الفصل الدراسي الثاني من العام الدراسي ٢٠٠٢ - ٢٠٠٣م ، والبالغ عددهم (٣٣٠٧) منهم (٤٧١ طلاب) و (٢٨٣٦ طالبات).

٣ - عينة الدراسة:

تكونت عينة الدراسة من ٢٥٠ طالب وطالبة من طلاب كلية التربية اختيرت عشوائياً من القسمين العلمي والأدبي للكلية. وتشكل هذه العينة ما نسبته ٧,٥٦% من مجتمع الدراسة حيث تكونت العينة من ١٣٦ طالبة و ١١٤ طالب من كلية التربية. والجدول التالي يوضح توزيع عينة الدراسة.

جدول (١)
عينة الدراسة موزعة حسب التخصص والجنس (طلاب / طالبات)

التخصص	الجنس / الشعبة	طالبات	طلاب	المجموع
علمي	**رياضيات**	٣٢	١٨	٥٠
	كيمياء	٢٣	٢٧	٥٠
	فيزياء	٢٥	٢٥	٥٠

أدبي	لغة عربية	٣٠	٢٠	٥٠
	لغة إنجليزية	٢٦	٢٤	٥٠
المجموع		١٣٦	١١٤	٢٥٠

أداة الدراسة:

اختبار اليقظة العقلية وهو الاختبار الأول من بطارية الاستعداد العقلي للمرحلة الثانوية والجامعية، من إعداد رمزية الغريب. حيث تعتبر هذه البطارية أكثر مقاييس الذكاء استخداماً في البحوث العربية، وفي تقدير استعدادات الطلاب بعد أن يصلوا إلى مرحلة لا بأس بها من النضج العقلي. والغرض من اختبار اليقظة العقلية قياس مدى قدرة الفرد على إدراك العلاقات فاليقظة العقلية تجعل الفرد لماحا للعلاقة، دقيقاً في تمييز الفروق، لذلك فإن الاختبار يعتبر مقياساً للذكاء العام. (الغريب، ١٩٦٣)

وصف المقياس:

يتكون الاختبار من (٢٢) بنداً يشتمل كل منها على ستة رسوم يمكن وضعها في متوالية منتظمة وذلك بتغيير مكان رسمين منها، ويطلب من الفرد أن يرتبها بعد أن يدرك العلاقة التي تربط بينها ويجد الرسمين المتبادلين. والاختبار من نوع اختبارات القوة Power Test فهو لا يعتمد على السرعة. وتكون الدرجة على كل بند من هذه البنود إما (واحد) في حالة الإجابة الصحيحة أو (صفر) في حالة الإجابة الخطأ.

و بذلك فقد توافرت شروط إجراء التحليلات الخاصة بنموذج "راش" في هذا الاختبار، حيث يفترض النموذج أن تكون مفردات الاختبار من النوع الذي يتطلب الإجابة بصح أو خطأ، ولا تعتمد إجابة أي مفردة على إجابة المفردات الأخرى، كما يفضل أن يكون الاختبار من نوع اختبارات القوة التي لا تلعب السرعة دوراً في الإجابة عن مفرداته، ولا تتأثر الإجابة بعامل التخمين، وأن تقيس جميع مفردات الاختبار سمة أو قدرة أحادية البعد. (علام، ١٩٨٧ : ٢٦)

تم تطبيق المقياس علـى عينـة استطلاعية من الطـلاب عددهم (١٠٠) ، وتـم جمع بيانات إجابات الطلاب على المقياس لحساب الثبات والصدق كالتالي:

١ـ الثبـات: لتقدير ثبـات الاختبـار علـى عينـة الدراسـة تـم استخدام أكثـر مـن طريقة وذلك كما يلي:

- **أ) طريقة ألفا لكرونباخ:** حيث بلغت قيمة معامل ثبـات الاختبـار ٠,٨٥ وهي قيمة مرتفعة.
- **ب) التجزئة التصفية:** حيث بلغت قيمة معامل الثبـات بطريقـة "سبيرمان براون" ٠,٩١ وبطريقة جيتمان ٠,٩٠ وهي قيم مرتفعة وذات دلالـة إحصائية.
- **ت) طريقـة كيـودر ريتـشارد سـون الـصورة (٢٠):** حيـث بلغـت قيمـة معامل الثبات بهذه الطريقة ٠,٨٨ وهي قيمة مرتفعة أيضاً.
- **ث) ثبات الاتساق الداخلي:**

حيث كانت جميع قيم معاملات الارتباط بين درجة كل مفردة والدرجـة الكليـة للمقياس دالة عند مستوى ٠,٠١ والجدول التالي يوضح ذلك.

جدول (٢)

قيم معامل ارتباط درجة كل مفردة بالدرجة الكلية لمقياس اليقظة العقلية ن = ١٠٠

رقم المفردة	معامل الارتباط بالدرجة الكلية	رقم المفردة	معامل الارتباط بالدرجة الكلية
١	٠,٢٩	١٢	٠,٤٥

الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة

٢	٠,٣١	١٣	٠,٨٢
٣	٠,٣٩	١٤	٠,٤٦
٤	٠,٤١	١٥	٠,٥٥
٥	٠,٥٣	١٦	٠,٧٣
٦	٠,٥١	١٧	٠,٣٢
٧	٠,٧١	١٨	٠,٢٧
٨	٠,٨٢	١٩	٠,٢٨
٩	٠,٤٩	٢٠	٠,٤٧
١٠	٠,٤٣	٢١	٠,٣٧
١١	٠,٨١	٢٢	٠,٥٧

ر ≥ ٠,٢٠٠ دالة عند ٠,٠٥

ر ≥ ٠,٢٦٠ دالة عند ٠,٠١

<u>٢ـ الصدق:</u> استخدمت الباحثة أكثر من طريقة لتقدير صدق الاختبار على النحو التالي:

١ ـ صدق المحك:

حيث بلغت قيمة معامل الارتباط بين درجات عينة الدراسة على هذا الاختبار ودرجاتهم على اختبار الذكاء المصور لأحمد زكي صالح (صالح، ١٩٧٨) (٠,٧٥) وهي قيمة دالة عند مستوى ٠,٠١ ، وكذلك بلغت قيمة معامل ارتباط بين درجات عينة الدراسة على مقياس اليقظة العقلية ودرجاتهم على مقياس الذكاء اللفظي للمرحلة الثانوية والجامعية من إعداد جابر عبد الحميد جابر، ومحمود أحمد عمر(جابر وعمر، ١٩٩٣) (٠,٨٠) وهي قيمة ذات دلالة إحصائية عند مستوى ٠,٠١.

<u>٢ ـ صدق المقارنة الطرفية:</u>

بعد الحصول على درجات عينة الدراسة الاستطلاعية وعددها (١٠٠) تم ترتيبها ترتيباً تنازلياً، ثم استخدم اختبار "ت" لقياس دلالة الفروق بين مرتفعي الدرجة على المقياس (٢٧% أعلى) ومنخفضي الدرجة على المقياس (٢٧% أدنى) وبينت نتائج اختبار "ت" وجود فروق ذات دلالة إحصائية بين متوسطي المجموعتين المرتفعة والمنخفضة وذلك عند مستوى ٠,٠١ وهي تعني قدرة المقياس على التمييز بين طرفي القدرة التي يقيسها. والجدول (٣) التالي يوضح ذلك.

جدول (٣)

اختبار "ت" لدلالة الفروق بين متوسطات درجات مرتفعي ومنخفضي الدرجة على مقياس اليقظة العقلية

	المتوسط	الانحراف المعياري	عدد أفراد العينة	قيمة "ت"
مرتفعي الدرجة أعلى ٢٧%	١٩	٠,٩٦	٢٧	٣٣,٣ دالة عند ٠,٠١
منخفضي الدرجة أدنى ٢٧%	١٠	١,٠١	٢٧	

مما سبق يتضح أن مقياس اليقظة العقلية صالح للتطبيق على عينة الدراسة حيث تمتع بدرجة مرتفعة من الثبات وكان على درجة مناسبة من الصدق، مما يطمئن إلى استخدام نتائجه والاعتماد عليها، من الوجهة الكلاسيكية أو التقليدية للقياس.

خطوات السير في الدراسة:

١- تم تطبيق أداة الدراسة بعد التحقق من صدقها وثباتها على عينة الدراسة.

٢- بعد الحصول على نتائج إجابة الطلاب لعينة الدراسة على أداة الدراسة تم إجراء العمليات التاليــة عليهــا:

أولاً:

١- تم تقدير معامل الثبات لهذه الدرجات باستخدام طريقة كرونباخ ألفا حيث إنها أفضل الطرق ، وبلغت قيمة معامل الثبات بهذه الطريقة ٠,٨٧ وهذه القيمة لا تختلف كثيراً عن قيمة معامل الثبات على العينة الاستطلاعية والتي بلغت ٠,٨٥.

٢- تم التعويض في المعادلة التالية للحصول على الدرجات الحقيقية المناظرة للدرجات الخام: ت = م + ر أ (س – م). حيث "ت" الدرجة الحقيقية، رأ معامل ثبات الاختبار، م متوسط الدرجات، س الدرجة الخام. (El Sayyad and StoKer, 1980 : 91)

وبذلك فقد تم الحصول على الدرجات الحقيقية المناظرة للدرجات الخام لعينة الدراسة (طلاب ـ طالبات ـ علمي ـ أدبي) باستخدام المدخل الكلاسيكي أو التقليدي في القياس.

ثانياً: للحصول على القدرات الحقيقية للأفراد مقدرة باللوجيت باستخدام نموذج راش قامت الباحثة بتحليل نتائج مفردات الاختبار (اليقظة العقلية) باستخدام نموذج راش معتمدة في ذلك على الطريقة التقريبية التي قدمها كوهين (Cohen, 1976) والتي عرفت باسم طريقة كوهين التقريبية Cohens Approximation التي تعرف الآن باسم PROX فيما بين الطرق الأخرى المختلفة. وهذه الطريقة اقترحها كل من "رايت ودوجلاس " (١٩٧٧) و "رايت وستون " (١٩٧٩) ورايت و ميد وبل (١٩٨٠)، كما أشار إليها علام (١٩٨٥: ١٠٧ ـ ١٠٨) (علام ١٩٨٦ : ١٣ ـ ١٣٨) ، كاظم (١٩٨٨: ٦٥ ـ ٦٧) كاظم (١٩٨٨ ب : ٣١). وتعد هذه الطريقة أبسط الطرق المستخدمة في تحديد أو تقدير بارامتري (معلمي)

نمـوذج "راش" أي تقـدير كـل مـن صـعوبة المفـردة (ص) وقـدرة الفـرد (ق) بوحـدة اللوجيت.

وتتميز طريقة بروكس PROX بإمكانية إجرائها بدون الاستعانة بالحاسب الآلي - وإنما يمكن أن تجرى باستخدام آلة حاسبة - والقيم التقديرية الناتجة عن هذه الطريقة تقترب من القيم التقديرية التي نحصل عليها باستخدام الطرق الأكثر تعقيداً والتي تعتمد على استخدام الحاسب الآلي مثل طريقة Ucon أو طريقة BICAL أو طريقة UFORM و غير ذلك من الطرق.

وتتلخص طريقة بروكس Prox لتقدير بارامتري نموذج راش في الخطوات التالية:

١ـ إعداد مصفوفة البيانات الخاصة بمفردات الاختبار التي تبين توزيع الدرجات التي حصل عليها كل فرد من أفراد العينة التي أجرى عليهم الاختبار في كل مفردة من مفردات الاختبار، وهذه الدرجات إما واحد صحيح إذا كانت الإجابة صحيحة أو صفر إذا كانت الإجابة خطأ.

٢ـ يستبعد الأفراد الذين أجابوا عن جميع المفردات إجابة صحيحة أو إجابة خاطئة، كما تستبعد المفردات التي أجاب عنها جميع أفراد العينة إجابة صحيحة أو إجابة خاطئة.

٣ـ ترتيب الدرجات الكلية التي حصل عليها جميع الأفراد في كل مفردة، وكذلك الدرجات الكلية لكل فرد في الاختبار ككل.

٤ـ تحول هذه المجموعات من الدرجات الكلية إلى نسبة مئوية.

٥ـ تحول هذه النسب المئوية إلى ترجيحات لوغارتمية Log Odds أو اللوجيت Logits وذلك بأخذ اللوغارتم الطبيعي Ln لخارج قسمة النسب المئوية للإجابات الخطأ على النسب المئوية للإجابات الصحيحة لكل مفردة. وكذلك أخذ اللوغاريتم الطبيعي لخارج قسمة النسب المئوية للأفراد الذين أجابوا إجابة صحيحة

على النسب المئوية للأفراد الذين أجابوا إجابة خطأ عن كل مفردة من مفردات الاختبار. وبذلك تتحول النسب المئوية التي تنحصر بين (صفر ، ١) إلى ميزان خطي Linear scale جديد للمتغير أو السمة الكامنة يمتد تدريجه من ∞- إلى +∞ وتزداد قيم هذا المتغير أو السمة بزيادة النسب المئوية للإجابات الخطأ في حالة "صعوبة المفردات" Item Difficulty ، بينما تزيد قيمته بزيادة النسبة المئوية للإجابات الصحيحة في حالة "قدرة الأفراد" Person Ability.

٦ــ يحسب متوسط وتباين كل توزيع من توزيعات هذه الترجيحات اللوغاريتمية التي تم الحصول عليها في الخطوة السابقة. ويعتبر متوسط الترجيح اللوغاريتمي للمفردة Mean Item Logit هو المركز الذي يناظر الصفر على ميزان تدريج المفردات Item calibration scale.

٧ــ يستخدم تباين كل من الترجيح اللوغاريتمي للمفردات Item logit والترجيح اللوغارتمي للأفراد Person Logit في حساب عاملي التعديل Expansion Factors أحدهما للمفردات والآخر للأفراد، ويستخدم هذان العاملان في حساب وتقدير قدرة كل فرد (ق) متحررة من أثر اختلاف صعوبة المفردات Test-Freed person Abilities . وكذلك في حساب وتقدير صعوبة كل مفردة (ص) من مفردات الاختبار متحررة من أثر اختلاف قدرة أفراد العينة.

وتكمن الحاجة إلى استخدام هذين العاملين لأن صعوبة المفردات تعتمد على تشتت القدرات كلما بدت المفردات متقاربة في صعوبتها، وكذلك كلما زاد تشتت صعوبة مفردات الاختبار كلما بدت قدرات الأفراد أكثر تقارباً. وهذه الآثار الناجمة عن تشتت قدرة أفراد العينة Sample Spread واختلاف صعوبة مفردات الاختبار يجب استبعادها من تقدير كل من صعوبة المفردات وقدرة الأفراد لكي تصبح هذه التقديرات متحررة من أثر خصائص العينة Sample Freed ومن أثر الاختبار Test

.Freed

٨ ـ تحسب الأخطاء المعيارية لهذه التقديرات لمعرفة مدى دقتها وبالطبع تقل هذه الأخطاء كلما كانت صعوبة مفردات الاختبار مناسبة لقدرات الأفراد. (يوسف: مرجع سابق ٤٥٨ ـ ٤٦٠) ، (علام: ١٩٨٥ : ١٠٧ ـ ١٠٨).

ثالثا: أصبح لدى الباحثة بعد الخطوات السابقة ثلاثة أنواع من الدرجات:

١ـ الدرجات الخام Obtained scores

٢ـ الدرجات الحقيقية المقدرة باستخدام النظرية الكلاسيكية True Score .

٣ـ السمات الكامنة أو قدرات الأفراد (الدرجة الحقيقية) مقدرة باللوجيت باستخدام نموذج راش.

رابعا: بالنسبة للإجابة عن السؤال الأول للدراسة استخدمت الدرجات الخام .Obtained Scores

خامسا: بالنسبة للإجابة عن السؤال الثاني للدراسة استخدمت قدرات الأفراد مقدرة بوحدة اللوجيت وهي القيم الناتجة من طريق بروكس PROX مباشرة لبارامتر قدرة الفرد ودون إجراء أي تحويلات عليها.

سادسا: بالنسبة للإجابة عن السؤال الرابع للدراسة قامت الباحث باستخدام التحويل الخطي التالي لقدرات الأفراد المقدرة باللوجيت : $\gamma = \alpha + \delta\chi$

حيث:

χ التدريج باللوجيت

γ التدريج الجديد

α عامل الموضع الذي يحدد وحدة القياس الجديدة.

δ عامل المسافة الذي يحدد وحدة القياس الجديدة.

ويستخدم هذا التحويل الخطي، للإبقاء على مميزات الوحدات المتساوية للوجيت المشتقة مباشرة من نموذج راش. وبذلك فإنه يمكن تحويل تدريج اللوجيت إلى تدريج مبني على معيار الجماعة له وحدات جماعية تسمى نيت (Nit) ويمكن تقدير قدرة الفرد B بهذه الوحدات الجديدة كما يلي :

$$B = \alpha + \delta(b \quad m)/s$$

حيث s , m هما المتوسط والانحراف المعياري مقدرة باللوجيت لقدرة الفرد المقدرة مباشرة من نموذج راش.

ويمكن اختيار قيم (s) بحيث تصبح وحدة القياس سهلة التذكر مثل ١٠ ، ٢٠ ، ٥٠ ، ١٠٠، كما يمكن اختيار قيم (α) بحيث يصبح متوسط عينة التقنين سهلة التذكر أيضاً.(Wright and Stone, 1979 : 198)

وقامت الباحثة بهذا التحويل الخطي لقدرات الأفراد المقدرة باللوجيت إلى وحدات النيت Nit حتى تكون المقارنة بين الدرجات الناتجة من معادلة " الصياد وستوكر " للدرجات الحقيقية المناظرة للدرجات الخام والدرجات الناتجة من نموذج راش مقدرة باللوجيت أكثر موضوعية في المقارنة وذلك لأن قيم اللوجيت أو تدريجات اللوجيت منخفضة (لا تتعدى ٥,٠٠ لوجيت) إذا ما قورنت بالدرجات الحقيقية ولذلك استخدم التحويل الخطي للقدرات المقدرة باللوجيت إلى قدرات مقدرة بالنيت.

سابعا: استخدم اختبار "ت" T-Test للتحقق من فروض الدراسة.

نتائج الدراسة وتفسيرها:

للتحقق من مدى انطباق نموذج راش على البيانات المستمدة من اختبار اليقظة العقلية المستخدم في هذه الدراسة، قامت الباحثة بتطبيق الخطوات السابقة لتحليل مصفوفة البيانات التي تشتمل على توزيع درجات طلاب العينة (٢٥٠) في

كل مفردة من مفردات الاختبار (٢٢ مفردة) وقد لاحظت الباحثة أنه لم يحصل أي طالب على واحد صحيح أو صفر في جميع المفردات، كما لا توجد أي مفردة أجاب عليها جميع الطلاب إجابة صحيحة أو خطأ، ولذلك استخدمت المصفوفة كاملة بعد ترتيبها، ودون استبعاد أي فرد من أفراد العينة، أو استبعاد أي مفردة من مفردات الاختبار.

وللحصول على القيم التقديرية لقدرات الطلاب اتبعت الخطوات السابقة والمحددة في طريقة PROX وذلك للحصول على السمات الكامنة المناظرة للدرجات الخام، واستخدام معادلة الدرجة الحقيقية التي تعتمد على ثبات الدرجات الخام ومتوسطها لتقدير الدرجات الحقيقية باستخدام المدخل الكلاسيكي، وتمت الإجابة عن أسئلة الدراسة على النحو التالي:

أولاً: الإجابة عن السؤال الأول: والذي نص على " هل تختلف الدرجات الخام obtained للطلاب على مقياس اليقظة العقلية باختلاف:

- الجنس (طلاب - طالبات).
- التخصص (علمي، أدبي)؟

وللإجابة عن هذا السؤال قامت الباحثة بالتحقق من صحة الفرضين الأول والثاني للدراسة وهما:

الفرض الأول: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للطلاب والطالبات في مقياس اليقظة العقلية.

الفرض الثاني: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للتخصصات الأدبية والتخصصات العلمية في مقياس اليقظة العقلية. والجدولين التاليين يوضحان نتائج ذلك.

جدول (٤)

اختبار "ت" للفرق بين متوسطي درجات الطلاب والطالبات على مقياس اليقظة العقلية

الدرجة الحقيقية المقدرة باستخدام نظرية السمات الكامنة

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
طالبات	١٧,٢٢	٤,٨٧	١٣٦	٤,٠٩٣	دالة عند ٠,٠٥
طلاب	١٥,٠٤	٣,١٩	١١٤		

يتضح من الجدول (٤) السابق:

١- وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للطلاب والطالبات على مقياس اليقظة العقلية لصالح الطالبات.

٢- وبذلك تم رفض الفرض الأول للدراسة.

جدول (٥)

اختبار "ت" للفرق بين متوسطي درجات التخصصات العلمية والأدبية على مقياس اليقظة العقلية

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
علمي	١٨,٨٥	٢,٠٥	١٥٠	١٥,٦٤	دالة عند ٠,٠٥
أدبي	١٥,٦٤	١,٩٦	١٠٠		

يتضح من الجدول (٥) السابق ما يلي:

١- وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الخام للتخصصات العلمية، والتخصصات الأدبية في مقياس اليقظة العقلية لصالح التخصصات العلمية.

٢- كذلك لم يتحقق الفرض الثاني للدراسة وتم رفضه.

فهل ستكون نتائج هذين الفرضين هي نفسها عندما تتم مقارنة متوسطي الطلاب والطالبات، ومتوسطي التخصصات العلمية والأدبية على مقياس اليقظة العقلية بعد تقدير الدرجات الحقيقية من النماذج الكلاسيكية، وتقدير قدرات الأفراد من نماذج السمات الكامنة (نموذج راش).

ثانياً: الإجابة عن السؤال الثاني: والذي نص على :هل تختلف الدرجات الحقيقية True score للطلاب على مقياس اليقظة العقلية والمقدرة باستخدام نظرية السمات الكامنة (نموذج راش) باختلاف:

- الجنس (طلاب، طالبات).
- التخصص (علمي، أدبي).

وللإجابة عن هذا السؤال قامت الباحثة باختبار صحة الفرضين الثالث والرابع وهما :

الفرض الثالث: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (المقدرة بنموذج راش) للطلاب والطالبات على مقياس اليقظة العقلية.

الفرض الرابع: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية (المقدرة بنموذج راش) للتخصصات العلمية والأدبية على مقياس اليقظة العقلية. والجدولين التاليين يوضحان ذلك.

جدول (٦)

اختبار "ت" للفرق بين متوسطي الدرجات الحقيقية (نموذج راش) للطلاب والطالبات على مقياس اليقظة العقلية

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
طالبات	٤,٠٥	٠,٧٥	١٣٦	٠,٦٦	غير دالة
طلاب	٤,٠٠	٠,٣٢	١١٤		

يتضح من الجدول (٦) السابق ما يلي:

١- عدم وجود فروق دالة بين متوسطي الدرجات الحقيقية للطلاب والطالبات على مقياس اليقظة العقلية.

٢- وبذلك فقد تحقق الفرض الثالث للدراسة وتم قبوله.

٣- وهذه النتيجة لهذا الفرض تختلف عن نتيجة نفس الفرض في حالة الدرجات الخام.

جدول (٧)

اختبار "ت" للفرق بين متوسطي الدرجات الحقيقية للتخصصات العلمية والأدبية على مقياس اليقظة العقلية

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
علمي	١,٩٩	٠,٣٩	١٥٠	٠,٨٠	غير دالة
أدبي	١,٩٥	٠,٤٠	١٠٠		

يتضح من الجدول (٧) السابق ما يلي:

١- عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية لكل من التخصصات الأدبية والعلمية.

٢- وبذلك فقد تحقق الفرض الرابع للدراسة ومن ثم تم قبوله.

٣- كذلك هذه النتيجة لهذا الفرض تختلف عن نتيجة الفرض الثاني للدراسة وهو نفس الفرض، ولكن على الدرجات الخام.

يتضح من نتائج إجابة السؤال الثاني للدراسة، عدم وجود فروق بين الطلاب والطالبات على مقياس اليقظة العقلية، وكذلك عدم وجود فروق دالة بين التخصصات العلمية والأدبية، وهذه النتيجة تختلف عن نتيجة الإجابة عن السؤال الأول والذي تم من خلاله التحقق من صحة نفس الفرضين للسؤال الثاني، ولكن باستخدام الدرجات الخام، حيث اتضح في الإجابة عن السؤال الأول بفرضيه وجود فروق دالة بين متوسطي الدرجات الخام للطلاب والطالبات على مقياس اليقظة العقلية لصالح الطالبات، ووجود فروق ذات دلالة إحصائية بين متوسطي درجات التخصصات العلمية والأدبية لصالح التخصصات العلمية. واختلاف نتائج الفرضين الأول والثاني عن نتائج الفرضين الثالث والرابع يؤدي إلى استنتاجات وتساؤلات كثيرة، حيث إنه في ضوء الإجابة عن هذين السؤالين الأول والثاني نحن في حاجة إلى إعادة جميع الدراسات والبحوث التي تم إجراؤها، والاعتماد في معالجتها على الدرجات الخام التي حصلنا عليها، حيث إنه من الواضح أن وجود الشوائب في هذه الدرجات تؤثر بصورة أو بأخرى على نتائج اختبار صحة الفروض باستخدام الأساليب الإحصائية المختلفة، وعليه فإن هذه النتائج التي تعتمد على الدرجات الخام في التحليلات الإحصائية نتائج مضللة ويجب إعادة النظر فيها ومراجعتها، ومحاولة استخدام الاتجاهات الحديثة (راش)، بل يجب استخدامها في كل دراسة للوصول إلى نتائج دقيقة وأكثر كفاءة وخالية من الأخطاء وذلك بالاعتماد على الدرجات الحقيقية أو

السمات الكامنة للأفراد بدلاً من الاعتماد على الدرجات الخام وما بها من شوائب.

كذلك فإن نتائج هذين السؤالين يفتح المجال أمام الباحثين لإجراء المزيد من البحوث التي تدعم هذه النتائج أو تضحدها.

ثالثاً: بالنسبة للإجابة عن السؤال الثالث للدراسة والذي نص على: هل تختلف الدرجات الحقيقية True scores للطلاب على مقياس اليقظة العقلية والمقدرة باستخدام النظرية الكلاسيكية باختلاف:

- الجنس.
- التخصص.

ولذلك قامت الباحثة بالتحقق من صحة الفرضين الخامس والسادس للدراسة وهما:

الفرض الخامس: لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية للطلاب والطالبات على مقياس اليقظة العقلية.

الفرض السادس : لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بالنظرية الكلاسيكية للتخصصات العلمية والتخصصات الأدبية في مقياس اليقظة العقلية. والجدولين التاليين يوضحان ذلك.

جدول (٨)

اختبار "ت" للفرق بين متوسطات الدرجات الحقيقية للطلاب والطالبات على مقياس اليقظة العقلية

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
طالبات	١٦,٠٢	٢,٢٦	١٣٦	٠,٠٧٥	غير دالة
طلاب	١٦,٠٠	١,٨٩	١١٤		

يتضح من الجدول (٨) السابق:

١- عدم وجود فروق ذات دلالة إحصائية بين متوسطي الطلاب والطالبات بالنسبة للدرجات الحقيقية على مقياس اليقظة العقلية.

٢ـ وبالتالي فقد تم قبول الفرض الخامس للدراسة والتحقق من صحته.

جدول (٩)

اختبار (ت) للفرق بين متوسطي الدرجات الحقيقية للتخصصات العلمية والتخصصات الأدبية على مقياس اليقظة العقلية

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
علمي	١٧,٢٤	١,٨٩	١٥٠	١,١٢	غير دالة
أدبي	١٦,٩٩	١,٤٥	١٠٠		

يتضح من الجدول (٩) السابق ما يلي:

١ـ عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية للتخصصات العلمية والتخصصات الأدبية في مقياس اليقظة العقلية.

٢ـ وبذلك فإن الفرض السادس للدراسة صحيح وتم التحقق منه وقبوله.

٣ـ تؤكد نتائج السؤال الثالث بفرضيه الخامس والسادس، نتائج السؤال الثاني للدراسة، وهو عدم وجود فروق بين الطلاب والطالبات من ناحية وبين التخصصات العلمية والأدبية من ناحية أخرى عندما تم معاجلة البيانات إحصائيا بالاعتماد على الدرجات الحقيقية المقدرة من النماذج التقليدية، على عكس ما توصلت إليه نتائج السؤال الأول والذي اعتمد التحليل الإحصائي له على الدرجات الخام للطلاب على مقياس اليقظة العقلية. وهذا يزيد من التأكيد على أننا في حاجة إلى المزيد من البحوث والدراسات التي تدعم نتائج هذه الدراسة، أو تخالفها.

رابعاً: بالنسبة للإجابة عن السؤال الرابع للدراسة والذي ينص على: هل تختلف الدرجات الحقيقية المقدرة باستخدام نموذج راش عن الدرجات الحقيقية المقدرة باستخدام النماذج الكلاسيكية؟

قامت الباحثة باختبار صحة الفرض السابع للدراسة الذي ينص على :

_**الفرض السابع:** لا توجد فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بنموذج راش وتلك المقدرة بالمدخل الكلاسيكي لمقياس اليقظة العقلية. والجدول التالي يوضح ذلك.

جدول (١٠)

اختبار (ت) للفرق بين متوسطي الدرجات الحقيقية المقدرة بنموذج راش والمقدرة بالمدخل الكلاسيكي لمقياس السعة العقلية

	المتوسط	الانحراف المعياري	عدد الأفراد	قيمة "ت"	مستوى الدلالة
الدرجات المقدرة بنموذج راش	١٦,٩٠	٥,١٢	٢٥٠	٠,٣٥	غير دالة
الدرجات المقدرة بالمدخل الكلاسيكي	١٦,١٥	٤,٣٨	٢٥٠		

يتضح من الجدول (١٠) السابق ما يلي:

١ـ عدم وجود فروق ذات دلالة إحصائية بين متوسطي الدرجات الحقيقية المقدرة بنموذج راش والمقدرة بالمدخل الكلاسيكي لمقياس اليقظة العقلية.

٢ـ تعتبر نتيجة هذا السؤال موضع للدهشة وإثارة كثير من التساؤلات على الرغم من أنها تتفق والإجابة على السؤالين الثاني والثالث من حيث اتفاق نتائجهما، وبالتالي يتوقع عدم وجود فروق ذات دلالة إحصائية بينهما، وهذا ما أكدته نتائج الإجابة عن هذا السؤال، والمثير للدهشة هو أن نماذج السمات الكامنة وكما سبق وأن بينا في الإطار النظري لهذه الدراسة جاءت لمعالجة كثير من القصور وأوجه النقد الموجه للمدخل الكلاسيكي في القياس، على الرغم من دقة هذه النماذج في تقدير الدرجات الحقيقية أو السمات الكامنة التي تقف وراء إجابات الأفراد على الأسئلة بالنسبة لأي اختبار، إلا أنها لم تختلف عن نتائج المدخل الكلاسيكي، والذي اعتمد على قيم معاملات الثبات في تقدير الدرجات الحقيقية.

توصيات الدراسة والبحوث المقترحة:

إن نتائج هذه الدراسة تلفت أنظار المختصين بالقياس والتقويم النفسي والتربوي إلى أهمية نماذج السمات الكامنة في تحليل نتائج البحوث والدراسات

السابقة، بالإضافة إلى عدم الاعتماد في التحليلات الإحصائية على الدرجات الخام فقط للمقاييس المستخدمة في أي دراسة، بل يجب أيضاً الاعتماد على نتائج نظرية السمات الكامنة أو الاعتماد على الدرجات الحقيقية للحصول على نتائج دقيقة وخالية من الأخطاء والشوائب، وبالتالي تكون قراراتنا صائبة ومبنية على أسس علمية سليمة.

كما أن نتائج هذه الدراسة تثير الكثير من الأسئلة والمقترحات لبحوث ودراسات أخرى، ومن هذه الدراسات المقترحة:

١ـ هل يمكن الحصول على نفس نتائج هذه الدراسة في حالة استخدام نموذج آخر غير نموذج راش أحادى البارامتر مثل نموذج لورد ثنائي البارامتر أو نموذج بيرنبوم ثلاثي البارامتر؟

٢ـ هل إذا قمنا بإعادة الدراسات والبحوث السابقة والتي تمت في جميع التخصصات وقمنا بتحليل نتائجها معتمدين على الدرجات الحقيقية بدلاً من الدرجات الخام، سوف نتوصل لنفس النتائج التي توصلت إليها هذه الدراسات في حالة استخدام الدرجات الخام في التحليلات الإحصائية؟

٣ـ هل تختلف نتائج هذه الدراسة في حالة استخدام اختبارات ومقاييس مختلفة على عينات مختلفة؟

المراجع

أولاً المراجع العربية:

١ـ أبو حطب، فؤاد ، وعثمان سيد أحمد، وصادق ، آمال (١٩٩٧). **التقويم النفسي**، ط٤ ، الأنجلو المصرية، القاهرة.

٢ـ الزيات، فتحي (١٩٨٩). أثر قوة تشتيت البدائل في أسئلة الاختيار من متعدد على صدق وثبات الاختبار، دراسة تحليلية، **مجلة كلية التربية، الجامعة المفتوحة، العدد ١١** ، ٨٧ ـ ١٠٨.

٣ـ السيد، فؤاد البهي (١٩٧٨). **علم النفس الإحصائي وقياس الفعل البشري**، دار الفكر العربي ، القاهرة.

٤ـ الشرقاوي، أنور محمد، كاظم أمين، الشيخ، سليمان الحضري (١٩٩٦). **اتجاهات**

معاصرة في القياس والتقويم النفسي والتربوي ، الأنجلو المصرية، القاهرة.
٥ـ الصياد، عبد العاطي أحمد (١٩٨٨). **نحو بناء مؤشر إحصائي جديد لتقدير الثبات بطريقة التطبيق وإعادة التطبيق لأداة القياس،** بحوث المؤتمر الرابع لعلم النفس في مصر، ص ص ٣٠٢-٣٢٤، مركز التنمية البشرية والمعلومات. القاهرة.
٦ـ الصياد، عبد العاطي أحمد (١٩٩٠). **المعامل ألفا لكرونباخ ونوع استجابة المستجيب لعبارة "اكتب أسمك إذا رغبت** ، بحوث المؤتمر السنوي السادس لعلم النفس في مصر، الجزء الثاني ، ص ص ٦٥٧ ـ ٦٩٦ ، الجمعية المصرية للدراسات النفسية، القاهرة.
٧ـ الغريب، رمزية (١٩٦٣). **بطارية الاستعداد العقلي للمرحلة الثانوية والجامعية،** دار النهضة العربية، القاهرة.
٨ـ القرشي، عبد الفتاح (١٩٨٦). اتجاهات جديدة في أساليب تقويم الطلاب، **مجلة رسالة الخليج العربي، ١٨**(٦)، ٣ ـ ٣٢.
٩ـ سعيد، محمد حسين (٢٠٠٠). **درجات امتحان الثانوية العامة: دراسة سيكومترية،** رسالة ماجستير، غير منشورة، كلية التربية، بني سويف، جامعة القاهرة.
١٠ـ صالح، أحمد زكي (١٩٧٨). **اختبار الذكاء المصور**، النهضة المصرية، القاهرة.
١١ـ جابر، عبد الحميد جابر، وعمر، محمود أحمد (١٩٩٣). **اختبار الذكاء اللفظي للمرحلة الثانوية والجامعية،** دار النهضة العربية، القاهرة.
١٢ـ عبد الرحمن، سعد (١٩٩٨). **القياس النفسي النظرية والتطبيق،** ط٣ ، دار الفكر العربي، القاهرة.
١٣ـ عبد السلام، نادية محمد (١٩٨٧). دراسة في حساب ثبات وصدق اختبار هدفي المرجع، **مجلة علم النفس، النهضة المصرية العامة للكتاب، ٢،** ١٥٥ ـ ١٦٥.
١٤ـ عبده، عبد الهادي السيد، عثمان، فاروق السيد (١٩٩٥). **الإحصاء التربوي والقياس النفسي**، دار المعارف، القاهرة.
١٥ـ علام، صلاح الدين محمود (١٩٨٢). استخدام نماذج السمات الكامنة في بناء بنوك الأسئلة، **صحيفة التربية، العدد ٢** (٣٣)، ٣٨ ـ ٤٤. القاهرة.
١٦ـ علام، صلاح الدين محمود (١٩٨٥). تحليل بيانات الاختبارات العقلية باستخدام نموذج راش اللوغاريتمي الاحتمالي "دراسة تجريبية"، **المجلة العربية للعلوم الإنسانية، ١٧**(٥)، ١٠٠ ـ ١٢٢، الكويت.
١٧ ـ علام، صلاح الدين محمود (١٩٨٦). **تطورات معاصرة في القياس النفسي**

والتربوي، جامعة الكويت، الكويت.

١٨- علام، صلاح الدين محمود (١٩٨٧). دراسة موازنة ناقدة لنماذج السمات الكامنة، والنماذج الكلاسيكية في القياس النفسي والتربوي، ، **المجلة العربية للعلوم الإنسانية، ٢٧** (٧). ١٨ - ٤٣، الكويت.

١٩- علام، صلاح الدين محمود (١٩٩٠). استخدام نموذج راش في بناء مقياس هدفي المرجع للمعارف الأساسية في إعداد خطة البحوث النفسية والتربوية ، **مجلة كلية التربية، جامعة الأزهر، ١٧**، ١٢٩ - ١٨٢، القاهرة.

٢٠- علام، صلاح الدين محمود (١٩٩٥). **الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية**، دار الفكر العربي، القاهرة.

٢١- عودة، أحمد وحوامدة، مفيد (١٩٩٦). خصائص توزيع الدرجات في المقررات الجامعية في ضوء نظام الدرجات بالتقديرات المعتمد بجامعة اليرموك ، **حولية كلية التربية، ١٣**، ٤٣٣ - ٤٥٦ ، جامعة قطر.

٢٢- غنيم، أحمد الرفاعي (١٩٩١). تعميم معامل ألفا لحساب معامل ثبات المقاييس ذات المفردات غير المتجانسة، **مجلة كلية التربية، ١٥**، ٢٠٧ - ٢٣٨ ، جامعة عين شمس.

٢٣- غنيم، أحمد الرفاعي (١٩٩٢). صيغ تقريبية لمعامل ألفا لتقدير معامل ثبات المقاييس ذات المفردات المتجانسة، **مجلة التربية، جامعة الأزهر، ٢٢**، ٣ - ٣٠.

٢٤- كاظم، أمينة محمد (١٩٨٨)أ. **دراسة نظرية نقدية حول القياس الموضوعي للسلوك (نموذج راش)**، مؤسسة الكويت للتقدم العلمي، الكويت.

٢٥- كاظم، أمينة محمد (١٩٨٨)ب. **استخدام نموذج راش في بناء اختبار تحصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج**، جامعة الكويت، الكويت.

٢٦- كاظم، أمينة محمد (١٩٨٨)ج. **مستوى العينة وتدريج بنك الأسئلة باستخدام نموذج راش، دراسة تجريبية**، القاهرة، المؤتمر الرابع عشر لعلم النفس في مصر، ص ص ٣٩٠ - ٤٢٦، الجمعية المصرية للدراسات النفسية.

٢٧- محمد، مجدي أحمد (١٩٩٦). **علم النفس التجريبي بين النظرية والتطبيق**، دار المعرفة الجامعية ، الإسكندرية.

٢٨- مراد ، صلاح أحمد ، و الشافعي، محمد منصور(١٩٩٨). أثر حجم العينة في دقة وكفاءة اختبارين في تدريج مشترك، القاهرة، **مجلة البحوث النفسية والتربوية، كلية التربية، جامعة المنوفية، ٢**(١٣)، ٩٧ - ١٤٨.

٢٩- يوسف، عماد عبد المسيح (١٩٩١). استخدام نموذج راش اللوغاريتمي - أحادي

د. غادة خالد عيــد
البارامتر في تحليل مفردات الاختبارات المعرفية مرجعية المعيار ثنائية القطب، **مجلة البحث فى التربية وعلم النفس، جامعة المنيا ، ٤(٤)**، ٤٤٣ ـ ٤٧٥.

ثانيا: المراجع الأجنبية :

30- Anthony, J. (1983). **Educational tests and measurement: An introduction**. New York: Harcourt Brace Jovanovich .

32-Caruso, C. (2002). Witkiewitz, Katie. Increasing the reliability of ability-achievement difference scores: An example using the Kaufman Assessment Battery for children. **Journal of Educational Measurement, 39** (1), 39-58.

33-El – Sayyad, A., & Stoker, H. (1980). Setting mastery test cut – score: An evaluative model approach. **Florida Journal Of Education Research, 22**, 71-107.

34-Garner, M., & Engelhard, G. (2002). An eigenvector method for estimating item parameters of the dichotomous and polytomous Rasch Models. **Journal of Applied Measurement, 3** (2), 107-28.

35-George, K. (1986). **Educational and psychological measurement**. New York: Collier Macmillan.

36-Hambleton, R., and Others (1978). Developments in latent trait theory. **Rev. of Educ. Res., 48** (4), 467-510

37-Hands, B., & Larkin, D. (2001). Using the Rasch Measurement Model to investigate the construct of motor ability in young children. **Journal of Applied Measurement, 2**(2), 101-20.

38-Joseph P., & Others (2001). Measuring disability: Application of the Rasch Model to activities of daily living (ADL/IADL). **Journal of Outcome Measurement, 4**(3), 681-705.

39-Komorita, G., & William K. (1965). Number F score points and the reliability of scales. **Educational And Psychological Measurement, 25**

(4), 987-995 .

40-Lord, F. (1980). **Application of item response theory to practical testing Problems**. N.J.: Laurence Erlbaum.

41- Lord, F., & Novick, M. (1968). **Statistical theories of mental test scores**. Reading, MA: Addison-Wesley.

42- Mathalon, D., Sullivan, E., Rawles J., & Pfefferbaum A. (1993). Correction for head size in brain imaging measurements. **Psychiatry Res: Neuroimaging, 50**, 121-139

43-Mehrens, W. (1984). **Measurement and evaluation in education and psychology**. Third Edition, New York: Holt, Rinehart and Winston.

44- Paul, J.(1991). Bayesian interpretation of test reliability. **Journal Of Educational And Psychological Measurement, 51**, 627-635.

45-Popham,W. (1980). **Criterion referenced measurement**. N.J.: Prentice Hall Inc.

46-Ronald, H., & Dato, N. (1983) Application of item response theory to criterion referenced test item selection. **Journal of Educational Measurement, 20**(4): 145-166.

47-Ronald, H., & Others (1983). Determining the length for criterion referenced test. **Journal of Educational Measurement, 20**(1): 169-181

48-Royce, S., & Eliot R. (1978). Does grade inflation decrease the reliability of grades ?. **Journal Of Educational Measurement, 1**(1): 37-41 .

49-Scharfe, E. (2002). Reliability and validity of an interview assessment of attachment representations in a clinical sample of adolescents. **Journal of Adolescent Research, 17**(5): 532-51.

50-Thomas O., & Ronald C. (2002). The reliability of test scores for the pervasive developmental disorders rating scale. **Psychology in the Schools, 39**(6): 605-11.

51-Waugh, F. (2001). Quality of student experiences at university : A Rasch measurement model analysis. **Australian Journal of Education, 45**(2):183-206.

52-William, W., & Stephen, G. (1985). **Educational measurement and testing**. London: Allyn And Bacon Inc.

53-Willmot, S., & Fowles, D. (1974). The **objective interpretation of test performance: The Rasch model applied**. England: National Foundation for Educational Research.

54-Wright, B., & Stone, M. (1979). **Best test design: A handbook For Rasch measurement**. Chicago: MESA Press.

55-Wright, B., & Mead, R. & Bell, S. (1980). BICAL: **Calibrating items with Rasch model**. Chicago: University of Chicago, Research Memorandum, No. 23.